AUUG Security Symposium 19-21 November 2001

Brisbane

Symposium Proceedings

http://www.auug.org.au/security2001/

AUUG Inc
PO Box 366, Kensington NSW 2033 Australia

Phone: 1-800 625 655 or +61-2-8824 9511
Fax: +61-2-8824 9522
Email: auug@auug.org.au

http://www.auug.org.au

ISBN: 0 9577532 3 3

UNIX® AND OPEN SYSTEMS USERS

Welcome from the symposium co-chairs

Welcome to the AUUG security symposium for 2001. The program assembled for you provides a great deal of variety and high quality content. The presenters come from a large range of roles and backgrounds and this represents the IT security industry in Australia today. Some people are system/network administrators, researchers, developers or managers. We thank all the speakers for their contribution to this symposium and to their industry.

The AUUG is rightly proud of its reputation as a group of highly skilled and practical technical people. This symposium and its program is part of that. Attendees can listen to and meet and debate issues with a large number of such people. This is in essence what the symposium is for - to provide such an opportunity to meet and debate current issues.

We would like to thank the tutorial presenters. This is the first time that the AUUG has trialed tutorials at the security symposium. We hope that this grows to become an annual opportunity for professionals to either update or diversify their knowledge.

Thanks must go to Liz Carroll from the AUUG office in Sydney. It is not easy to organise an event interstate especially during the preparation for the main AUUG national conference. Liz was a great help in organising the event and at all times was a very knowledgeable and helpful person with which to work - thank you Liz.

And finally - thanks to you for your support in attending and participating in this security symposium

Gary Gaskell
Warren Toomey

Table of Contents

KEY NOTE ADDRESS

Global Trends in Computer Security

By

Adrian McCullagh

FREEHILLS

Level 38
Central Plaza 1
345 Queen Street
Brisbane 4000

email: Adrian_mccullagh@freehills.com.au

Tel: 07 3258 6603
Fax: 07 3257 6444
www.freehills.com.au

8 August 2001

Introduction

Firstly, I would like to thank the organisers of this conference for the opportunity to speak to you today about "Global Trends in Security".

The economic reality is that business and government are more and more aligning themselves to transacting via the internet. This global migration to an open and obviously insecure communication infrastructure causes its own set of problems. On the one hand there are issues concerning payment and service delivery logistics, and on the other hand there are the security requirements of the organisation. As IBM has recently stated in their global advertising campaign "solutions for a small planet", which identifies the breakdown of the tyranny of distance through the global information infrastructure but this breakdown causes a number is security issues that need to be addressed by Information Security Officers (ISOs).

As a result of this communications/information/business revolution, there have been numerous reports published globally concerning security infractions. As this migration increases, it is likely that security will/has become the major focal point for management. That is, in the same manner that Boards of Directors will discuss strategies for the future direction of a business, these same people will also discuss the strategic security model for the business. Aligned to this will be the increased appointment of IT/Security experienced people to the Board of Directors of companies. The positioning of security is really a never-ending strategy that ensures the future protection of the organisation through the protection of its information assets.

The main theme that I want to emphasise is that society is firmly placed in the information era and therefore management has the obligation and fiduciary duty to protect all assets of the organisation, which must include all information assets. In addition to the security of information assets there arises the strategic positioning of the organisation to market and contract via the internet and where possible provide online fulfilment.

In this presentation I will briefly explain my thoughts on security, but more importantly I will concentrate on market risks and market values. This discussion will emphasis the global positioning currently taking place by various industry sectors such as the finance sector, health sector and government sector. Each sector has a different though complementary risk profile and therefore a different market value in security.

1.1 What is Computer Security

If we accept the proposition that commerce is utilising the Web as a delivery/marketing channel and that this infrastructure is open and insecure then Web security is the weakest link.

According to Garfinkel & Spafford, web security is:

> *"A set of procedures, practices and technologies for protecting Web servers, Web users, and their surrounding organisations.* ***Security protects*** *you* ***(the user)*** *against unexpected behaviour"*

The first thing to note about the above definition is that it does not solely concentrate upon the implementation of technology. Computer security involves a number of components, a major component of which includes the human interface namely procedures and practices. The technology used for computer security is complex, which is usually configured through software switches. The real task is to ensure that the various configured software switches DO NOT create inadvertently a conflict causing unintended security vulnerabilities.

There have in the past been a number of security policies and handbooks developed that assist ISOs in their deliberations concerning what is best for their relevant organisation.

There have been many security texts that have identified that security really revolves around three components, namely:

(a) What you know;

(b) What you have; and

(c) What you are.

WHAT You Know

This issue concerns the use of passwords or pass-phrases. Traditionally, this type of security mechanism has had its security questioned due to the inherent lack of appropriately of secure names. Passwords and pass-phrases are usually vulnerable to dictionary attacks and to social engineering attacks. As humans we have a substantial inherent laziness when it comes to developing appropriate passwords for system access. In order to remember the password or pass-phrase the individually usually opts for a common name or a name that may be specific to the relevant person is also easy to remember; hence the social engineering perspective and the social inquisitor at parties or other social events.

What You Have

This really relies upon some form of token technology. In order to gain access the token must also be present. The token may also (usually the case) require a activation code. Therefore to gain unauthorised access the perpetrator needs to two things namely:

(a) the token for access

(b) the password/pass-phrase for activation.

This makes unauthorised access more difficult but not insurmountable.

What You Are

Biometrics access originated in very high assurance areas (Defence). Recently, it has progressed to the financial sector and in some circumstances the health sector.

Again, biometrics has been aligned to token driven access control mechanisms.

New technology drivers

One technology that has recently been flavour of the month is Public Key Infrastructure technology. Instead of using pass-phrase technology PKI uses the security of "what you have" as the access

mechanism, it combines the security elements of what you know with what you have. The "what you have component is the private access key that is used to identify the access applicant.

PKI technology can assist in web security through the use of identity certificates and attribute certificates. It is possible though there a few applications exhibiting this feature to have a single sign mechanism. This is usually achieved through attribute certificate but can be achieved through other mechanism. If Public key technology is to be used for access control it is not necessary for a Trusted Third Party. A trusted directory is all that is needed.

But PKI deployment is not for the faint hearted. It is a relatively new commercial product though the underlying technology has been around since the late seventies with the public development of the RSA algorithm.

What I mean by faint hearted is that if a PKI is deployed it must be done correctly and as such it is not inexpensive. But in the high assurance markets the value of the organisation and the push to market share substantially outweighs the cost of deployment. In some industry sectors, such as health, banking, insurance, government and gaming, the market participants are in effect dictating the move to PKI.

1.2 Why have security

Why have computer security?

Let me tell you a little story, which concerns an organisation whose only asset was the information stored on its computers. This will explain why computer security is important and that it is not solely reliant upon the technology and is highly dependent upon the practices and procedures implemented.

This organisation was in the business of back loading. Back loading involves the ability to keep track of long distance haulage requirements and to arrange for empty trucks to be loaded after they have delivered their

goods to the relevant destination. The organisation was solely operating in the transport logistics market. That is, if truck A was delivering goods from San Francisco to New York and did not have a return consignment organised, the back loader will organise a back order for truck A either to San Francisco or to some location close to San Francisco.

The only asset of value to the back-order organisation was the information about location of trucks in the system, the clients who required the delivery of goods and whether they could match trucks to suppliers to delivery locations. The back order organisation did not own the trucks. It simply was in the business of matching empty trucks to client requirements for the fee paid by the clients.

One day, a disgruntled employee (was about to be dismissed) let loose a trojan horse that literally destroyed all the information on the system. The disgruntled also was able to tamper with the back up copies of the data. Now the perpetrator was caught and was jailed but this story does not involve the criminal action but concerns the civil suite that followed.

The shareholders of the back-order company brought a class action against the management of the company for breach of fiduciary duty in not adequately protecting the primary assets of the company. They succeeded in their claim, which was later in part settled under an insurance policy.

The purpose of this story is to show that a breach of security can have a devastating effect upon management. Consequently, security is important and more importantly, the court in this case took specific notice that modern society and in particular business existed within the information era and as such information is a highly valuable commodity that required protection. Further it was senior managements responsibility and fiduciary duty to make sure that all assets include information assets are adequately protected.

Incidentally the company went out of business and that is a true case in the USA.

What can be learnt from this case?

Firstly, the court took judicial notice that society and commerce being a sector of society exists in the information age.

Secondly, management had a substantial duty to protect all assets of the organisation and not just the physical assets. In particular, it included the information assets of the organisation.

Thirdly, management had breached the fiduciary duty to implement appropriate security and risk management procedures for the benefit of the organisation as a whole.

Finally, the shareholders had a direct cause of action against the management of the organisation and thus management had exposed their own assets in the action.

Hence, security has developed into a major focus for business and this will only become more pertinent as business develops in the online environment.

1.3 PKI - its role in a Security Framework

PKI can play an important part in a security framework but that is not its only attribute. One of the advantages of PKI is that if properly implemented it can provide a non-repudiation service.

According to McCullagh, Caelli and Little:

> *"Non-repudiation is the ability to prevent a party to a transaction from successfully denying that the relevant digital signature attached to an electronic transaction is not theirs."*

This involves the use of trusted systems and having evidence to support the claim that only the alleged signatory could have digitally signed the relevant transaction.

The issue of NON-REPUDIATION a real value proposition in the PKI provided, but this necessitates that it was deployed in a high assurance manner. High assurance primarily involves the use of smart tokens and thus two elements of security are utilised.

One of the main benefits of a PKI is that it is flexible to cover many situations from access control to authenticating the parties to a remote transaction. Within an access control market segment a PKI that uses attribute certificates greatly reduces administrative costs by substituting the access control list with a set of attributes that will dictate the control of

access. The person/entity seeking access will also possess an attribute certificate that will be compared to the access attributes. If they match or exceed the minimum requirements then access is granted otherwise access is denied.

So a PKI can simplify computer security as well as provide value in meeting or exceeding the legal requirements for evidentiary purposes in actions for non-repudiation.

Therefore, PKI technology can provide a number of elements that enhance security with access control as well as provide other benefits such as non-repudiation, authentication, and confidentiality. But this will come at a cost and as such before embarking on a PKI deployment there MUST be a business case to support the capital cost and the ongoing costs. In the next section I will discuss the important issues of Market Risk and Market Value.

2 Market Risk

2.1 What is Market Risk

All commercial organisations fit with a market, whether it is banking and finance or insurance or steel production, or education or just selling widgets. Being a commercial organisation, the management has an obligation either financial, fiduciary, or morally and legally (in some cases) to maximise the return on investment for the organisation. Therefore, the organisation must exist within a market so as to maximise its return on investment.

Being in a market there are environmental forces that influence the decisions made by management and the direction of the organisation. Some of these forces include regulatory frameworks, market perception, competitor involvement, and client requirements. For example, the banking sector has for decades taken the issue of risk management very seriously. The health sector has had a mixed view of risk management. If the risk involved human safety then the health sector was diligent in managing the relevant risk but if the risk involved financial/information risk the health sector has had a less than adequate approach. This is rapidly changing due to changes in regulation such as privacy legislation and financial reporting requirements.

Therefore, generally there is in the PKI environment a global trend that regulatory and market forces are requiring organisations to deploy PKI technology. Part of the market risk is what is the down side if the organisation does not deploy the technology. Will market competitors get a jump on the field and take not only a market share but also a mind share. But do not be fooled into thinking that mind share always results in increased benefits for the organisation. As will be discussed in the next section there must be a market value in the deployment of any new innovation.

2.2 The Market Value – Where does it sit

In the diffusion of any innovation, there must be a value proposition otherwise the market will not adopt the technology. The value proposition could be cost savings, or increased revenue or market perception. It appears to me that market perception has and continues to play an important part in the uptake of PKI. Even though PKI has certain benefits, it is not the panacea for all situations. I wish it was but reality dictates otherwise. The cost of implementing a PKI causes great concern for system administrators.

Before any organisation allocates substantial resources to any undertaking, there must be a sufficient business case that can support the expenditure life cycle. Some organisations like SPYRUS Inc, have developed through a lot of pain a business case methodology that will use to assist clients to determine whether there does exist a business case of the deployment of a PKI. For example, the clear advantages of deploying a PKI are:

(a) Authentication;

(b) Non-repudiation;

(c) Message integrity;

(d) Confidentiality.

Therefore, the benefits of deploying a PKI must outweigh the cost of such deployment. Coupled with this is the risk involved in project deployment. Will it (the Project) succeed by delivering the stated goals?

It is becoming clear globally that the hype of PKI is being replaced with a realistic position of where is the business case to support the expenditure lifecycle. For example, the Florida Department of Law Enforcement deployed successfully a PKI so that there could be a single sign on for in excess of 60 different databases for the police, sheriffs department and emergency workers. The business case was initially difficult but after some time the department realised that the deployment of a PKI could vastly reduce administrative costs in accessing multiple databases that the various county and state police needed.

2.3 Driving Forces – Who, What, When, Where & How

There are a number of driving forces influencing the deployment of PKI globally.

2.3.a Accreditation Schemes

Structures like "Gatekeeper" that promotes trust within the online framework are beneficial and are clearly a driving force. Globally, a number of countries are reviewing a gatekeeper scheme to see if it is applicable for them. But in doing so the issue that arises is whether the Gatekeeper bar has been set too high for the benefit obtained in getting gatekeeper accreditation. It is a good thing to have an accreditation scheme but this does come at a cost and as such an accredited supplier will pass on those costs to their respective markets. The cost of obtaining accreditation is not cheap and in reviewing those parties that have received accreditation there does not appear to be consistency by the accrediting authority. Notwithstanding this, there are clear benefits in the development of an accreditation scheme, and as such Australia should be congratulated.

2.3.b High Assurance Markets

Another driving force is the high assurance market that is demanding a better method to authenticate parties remotely. Aligned with this is the demand by the same market segment to have non-repudiation mechanisms. But the cost of non-repudiation is not inexpensive and as such there must be a business case to support this mechanism. In order to obtain true non-

repudiation, the environment must be trusted and this is only possible through the deployment of smart tokens operating on a trusted platform. The IDENTRUS Project is a clear example of this that is having global impact.

All of us may not like Banks but through our actions in depositing our savings with them, we certainly as a general statement trust them. Consequently, from a trust perspective the banking industry has a substantial advantage over other players in developing a global trust network. This is a global trend, which in Australia is exhibited by project Angus. I understand that the 4 major banks in Australia are the sole players in project Angus, which intends to issue IDENTRUS, accredited certificates.

Therefore, large international players are enticing various domestic players to join large clubs so that there is global coverage for the club. This is also being exhibited in the mining industry through the global buying clubs, through VORTALS. These clubs are expected to dominate the B2B market and require non-repudiation and authentication for support.

2.3.c Business Case

There is a global trend in the PKI market, which is becoming increasingly sophisticated with the technology to insist upon a business case to support the relevant project. Without a proper business case to support the deployment of a PKI, the project is doomed for failure. The business case MUST be honest and identify the rationale for the deployment of a PKI. Further it must identify alternatives and cost all possible solutions and then compare the various benefits and costs of each solution set. Therefore, business case development is a definite driving force.

2.3.d Legislation

In the USA, the HIPAA (Health Insurance Portability and Accountability Act 1996) has had a profound affect on the deployment of PKI. This legislation covers a myriad of Health issues including privacy. On December 21 of this year, the Privacy Amendment Act 2000 will come into force, which is Australia's compliance mechanism to the European Directive

on Privacy. The ambit of this Act is to cover all businesses whose gross turnover exceeds 3 million dollars and all health providers irrespective of turnover. Health information is a driving force for the deployment of PKI. For policy reasons health information falls within the high assurance area. National privacy principal 4 specifically provides that if an organisation collects personal information then that same organisation must also implement reasonable security for the protection of the collected information. What is "reasonable" is determined on a case-by-case basis. This provision really throws open the liability of protecting information in the modern era onto computer security professionals.

The enactment in a number of jurisdictions of electronic signature legislation has helped but there appears to have been a real difference in opinion between the framers of the laws and the business side of government when it comes to electronic signatures. The lawmakers do not want to be seen as endorsing any particular technology and as such there has been a real trend globally for technology neutral language. On the other hand, when it comes to the service delivery by government, PKI has been the only security technology endorsed. Therefore, whether governments like it or not they are by their own deployment strategies influencing the deployment of specific technology namely PKI.

I suggest that lawmakers will need to rethink their approach to electronic signature legislation and will need to enact technology specific legislation. If they do not, then I predict a substantial increase in identity fraud cases which if they could have reduced the incidence of through technology specific legislation.

2.4 Market Impediments – They are out there, some real - some imaginary

One of the key impediments to the diffusion of PKI is the perceived position that the solution is technical, where in fact it is really a business solution. Yes PKI does involve technology but no than a motor vehicle. It is the use of the technology that is supported by a sufficient business case that is important. The technology from an abstract level is relatively easy to

understand, what is uncertain is understanding the business requirements needed. As I have already stated, a PKI involves technology, process and procedures. The human intervention needs to be better understood.

Globally, there appears to be a trend that education in business processes and the benefits of PKI are starting arise. That is, not until there is a basic education program for business managers about the fundamentals of PKI will we see an increased uptake of PKI.

3 A Management not a Technology Issue

The technology that supports a PKI is now very stable and has been subjected to the "40 cycle washing machine" test. That is, the underlying technology has been subjected to quite substantial robustness tests by third parties. For example, third parties have tested SPYRUS technology and it has been accredited, for Common Criteria EAL 4 or ITSEC E3. Further the common algorithms have been subjected to extensive academic analysis to ensure their robustness and security.

Notwithstanding this, the ownership of a PKI is not for the technologists but must be owned by management. Hence, management must understand the business case. Management must be involved in the implementation of security policies and any business case to support the security policy. Without their involvement there will not be the appropriate adoption by the organisation or the belief in the necessary security policy by all parties.

4 The Crystal Ball – What the Future Holds

The issue of security is only going to become more focused. Computer suppliers will need to address this issue from the ground up and not has been the case in the past use a patchwork mentality. I appreciate that the original operating systems for PC work never intended to be used on networks or for that matter in commercial environments. But enough is enough, the major manufacturers have to take responsibility and start to address the issue of computer security from the ground up. The time is right.

Mobile computing is in part where the future lies. This technology is still in its infancy and therefore will not just yet take as big a slice of the market are

some analyse have advocated. As this technology matures you will see a migration of PKI technology to the mobile framework. Initially for the mobile workforce but eventually extending to the general consumer.

If one takes the timeframe for the deployment on PKI in a dynamic environment, I suppose the mobile computing will not achieve maturity in the market place for some 7 to 10 years. This could be shorter if the so-called "killer app" is developed. But this app will also need the killer market so that the uptake of the technology is increased.

It is my belief that the so-called killer market for mobile PKI will involve a combination of PKI, combination smart token (contact and contact less) and G3 PDA's. The market will involve banking with some sort of high assurance environment with lots of money involved at one end and extreme low assurance at the other end. I know I am being very vague but we at SPYRUS will shortly be working on a project that meets this characteristic in South-East Asia

I also believe that PKI will become ubiquitous but it will be hidden in the background and will become the norm when the general public becomes accustomed to signing electronic documents. This will take time. Nothing in this area happens overnight and changing a society's pattern of effecting commercial transactions is going to take time.

5 Conclusion

I hope in the short time that I have been talking that I have conveyed 6 important points as regards to the global trends that I see in PKI, namely:

(a) PKI will eventually become ubiquitous but in many respects hidden from the general business activity;

(b) Before an organisation deploys a PKI it must build a business case that can support not only the initial cost of the PKI but also support the ongoing expenditure;

(c) The opinion leaders in the PKI market are the finance sector, governments and the health sector and as such these sectors will drive the market for PKI deployment;

(d) PKI currently operates in the high assurance market and this will continue for the foreseeable future;

(e) PKI will migrate to mobile computing but this will take sometime as this technology matures.

(f) Nothing is easy otherwise everyone would be doing it but with PKI care and experience is the solution.

I thank you the time to discuss my views on global trends within PKI.

ARE THERE ANY QUESTIONS

New Trends in Authenticating Payments On-line

Duncan Unwin
Director of Marketing
Nov 2001

QSIPayments

- Background to on-line payments
- The problem of Fraud
- The traditional solution – SET
- Problems with SET
- The replacements
  - VISA's 3-D Secure
  - MasterCard's SPA
  - A comparison between schemes
- Financial Institution Considerations
- Merchant Considerations
- Cardholder Consideration
- Summary

How credit card payments work on-line today

5123 4567 8901 2345 Exp 12/02

Person shopping on-line enters credit card number on line into merchant's web form

The merchant sends an authorization request via POS-type interface to their Acquiring Bank

The Issuers authorizes the transaction as a 'Card not present' transaction

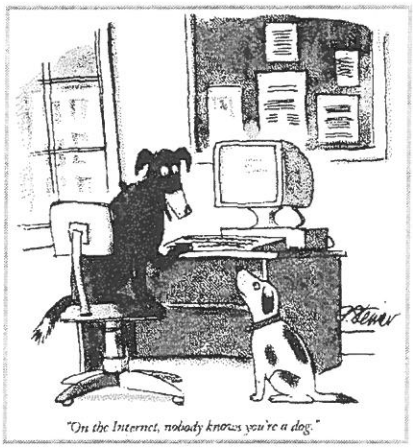
QSI Payments

3

- Card Not Present
  - Payment card is never proven to be in the possession of purchaser
  - Payment card transaction is never provably authorized by card holder
- First Person Fraud
  - When the owner of the card fraudulently claims not to have made purchase
- Lost and Stolen Fraud
  - When a card or card number is stolen and used by a person other than the card holder

QSI Payments

4

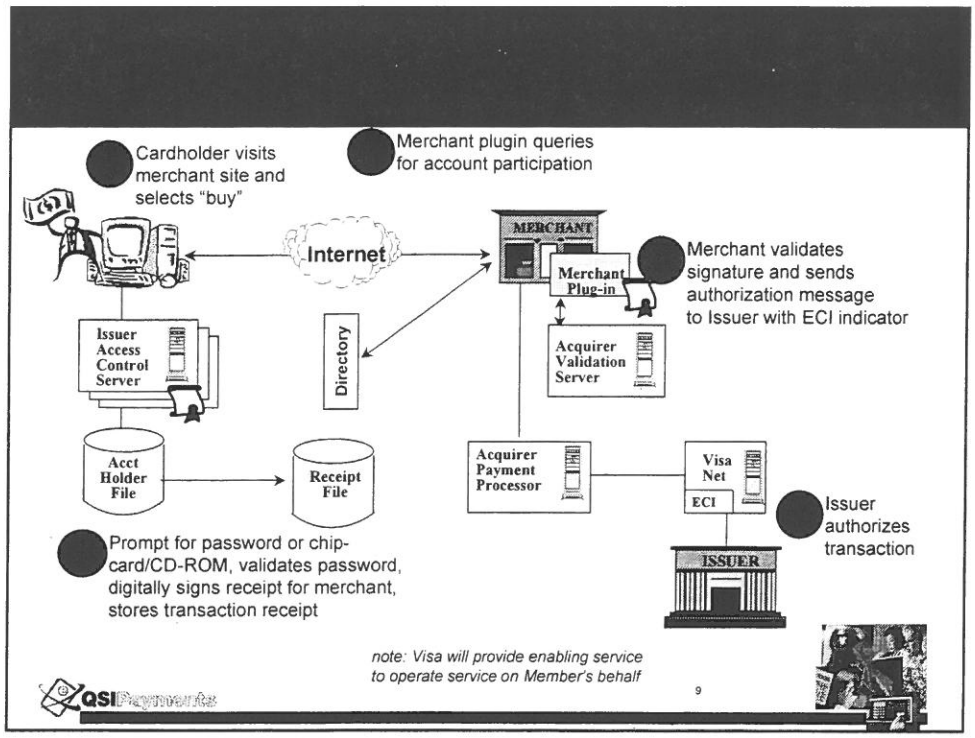


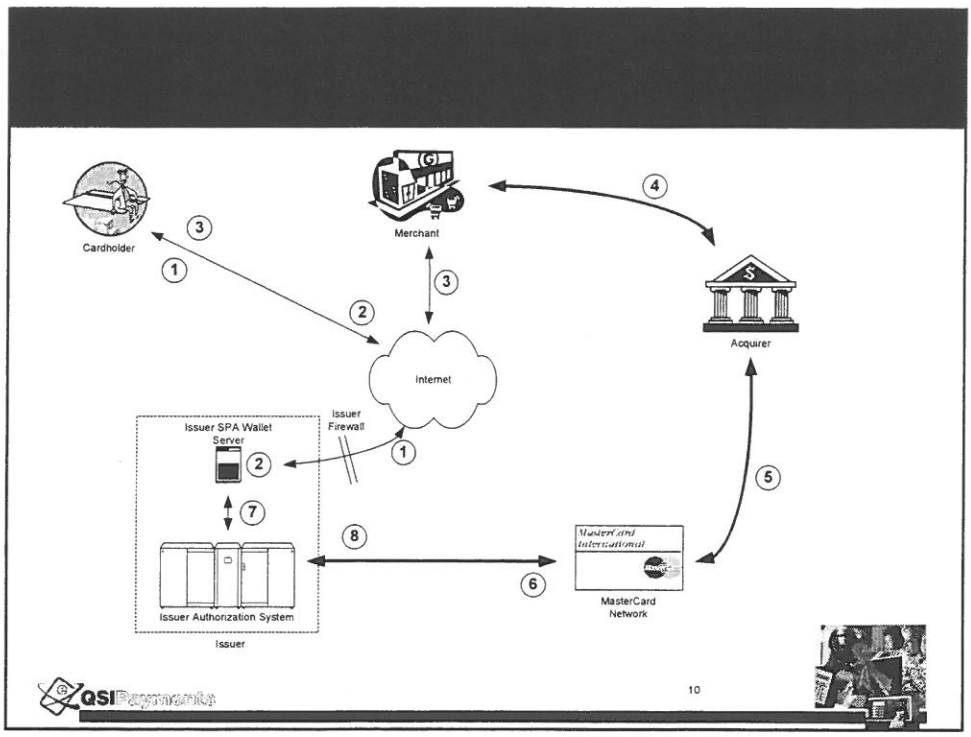
"On the Internet, nobody knows you're a dog."

- Traditional Management Approaches
  - Address Verification Service (AVS)
  - CSC2, CVV2, CID
  - Protection from theft at Merchant and In-transit
    - Card details via SSL to Bank not Merchant
  - Fraud screening and risk scoring
    - eFalcon, Trustmarque, Cybersource

- Uses PKI, Card holder SET Wallets, Merchants have SET POS
- Good points
  - Security strong
- Bad Points
  - Deployment difficult
    - Merchant POS has to set up
    - Cardholders need to enrol and install software
  - Value proposition weak
    - For Merchant
    - For Cardholder
  - Slow and not reliable in when released

- VISA's 3-D Secure (3-DS) aka the Verified by Visa programme
- MCI's Secure Payer Authentication (SPA)
- AMEX ??? but probably EMV smart card in long term
- Both MCI and VISA schemes work by getting cardholders to register with their Issuer's Authentication tool (server wallet or access control server)
- Some talk of convergence of 3-DS and SPA but unlikely in the near term (IMHO)
- Merchants need to support both!


Cardholder visits merchant site and selects "buy"
Merchant plugin queries for account participation
Internet
MERCHANT
Merchant Plug-in
Merchant validates signature and sends authorization message to Issuer with ECI indicator
Issuer Access Control Server
Directory
Acquirer Validation Server
Acct Holder File
Receipt File
Acquirer Payment Processor
Visa Net
ECI
Issuer authorizes transaction
ISSUER
Prompt for password or chip-card/CD-ROM, validates password, digitally signs receipt for merchant, stores transaction receipt
note: Visa will provide enabling service to operate service on Member's behalf
QSI Payments
9


Cardholder
Merchant
Acquirer
Internet
Issuer SPA Wallet Server
Issuer Firewall
Issuer Authorization System
Issuer
MasterCard Network
QSI Payments
10

Attribute	VISA 3-D Secure	MCI SPA
Non-repudiation	Yes	Yes
Confidentiality	No	No
Integrity	Partial	No
Easy to deploy – Card holders	Yes	No (requires wallet)
Easy to deploy - Merchant	No (requires MPI)	Yes

- Issuing Systems
  - Need to have Server Wallet and/or Access Control Server
  - Integration into Issuing Systems
  - Switch updates to support UCAF and ECI
- Card holder deployment and education
- Shift of risks from Acquirer to Issuer
- Acquiring Systems
  - Switch changes
- Merchant support and education

- Need to support BOTH standards
- May be forced to support by Acquirer
- Possible cost savings from Fraud
- What to do with non-authenticated transactions
- Timeframe for implementation

- Benefits / Risks of card present transactions
  - Q: What does 'liability shift' mean?
  - A: Even if you do nothing, it may be treated as a card present transaction if merchant supports the new standards.

- *New Standards for on-line authentication of card payments*
- *VISA 3-D Secure (Verified by Visa)*
- *MasterCard SPA*
- *Merchants, Banks and Card Holders need to be aware of developments*
- *Can reduce cost of on-line fraud*

Thank you

Contact:

Duncan.Unwin@qsipayments.com

Ph: +61-7-3210 2522

Advanced log analysis.

See http://www.auug.org.au/security2001

Information Security, the Australian Privacy Regime, and What It Means for IT Security Practioners

Brian Denehy & Bernard Hill
(b.denehy@90east.com)
90East

First, a brief background to the Privacy Act.

The Privacy Act was passed in 1988, and came into effect in 1989. It was largely designed to compliment the Australia Card, by giving reassurance that information collected under the Card would be protected. However, as you would know, the Australia Card did not survive a voter backlash, but the Privacy Act did. It applied to Federal Government agencies, and a few other private sector organisations, until December last year when the Federal Parliament passed a bill to extend the Act to the private sector, commencing in December this year. The year delay was to give business time to comply with the Act.

Before I move on to share with you the particular challenges that business will face in implementing the Act, there are two points that need to be understood.

First, the Act applies to 'personal information'. What is personal information? It is information *about* an individual. It needs to be noted that it can be about a person, and therefore does not have to have been collected *from* me, but could have been provided by a third party, and I may not even know that it is held. It can also include an opinion, whether true or false, and the significance of this I will explain in a moment.

So it is information about an individual from which that person's identity can reasonably be ascertained.

The second point that needs to be noted is that the Act is technology neutral. It applies to personal information no matter what medium in which it is stored, whether it be in ones and zeros in the digital medium, or on audio tape or on pieces of paper in filing cabinets. There has been a great deal of emphasis on the Act's application to the IT medium, with discussion on cookies, browser tracking etc. However, the collection and storage of personal information on other media should not be overlooked.

I'd like to discuss the area of the Act that I believe will be particularly challenging to business, and that is the obligation to protect personal information against 'misuse and loss and from unauthorised access, modification or disclosure' contained in National Privacy Principle 4.1. NPP 4 is one of ten National Privacy Principles

contained in the Act. The NPPs detail the rights and obligations regarding securing personal information. NPP 4.1 states:

> 4 Data security
>
> 4.1 An organisation must take reasonable steps to protect the personal information it holds from misuse and loss and from unauthorised access, modification or disclosure.

In addition to the bare legislation, the Office of the Federal Privacy Commissioner (OFPC) has issued Guidelines that indicate some factors the Commissioner may take into account when handling a complaint. The guidelines are advisory only and not legally binding.

The key phrase in NPP 4 is 'reasonable steps'. This term is common in law, and the question of what is reasonable is ultimately a legal, and not a technical one. However, the Commissioner or a tribunal would obviously hear expert evidence on the matter before making a decision whether reasonable steps were taken.

According to the Guidelines, what are reasonable steps to secure personal information 'will depend on the organisation's particular circumstances'. The Guidelines list a number of factors that could be taken into consideration, being:

- the sensitivity of the personal information the organisation holds;
- the harm that is likely to result to people if there is a breach of security;
- how the organisation stores, processes and transmits the personal information (for example, paper-based or electronic records);
- the size of the organisation (the larger the organisation, the greater the level of security likely to be needed).

The Guidelines then provide 'Tips for compliance' and list a number of steps that an organisation could take to comply with NPP 4, which include:

- risk assessment - identifying the security risks to personal information held by the organisation and the consequences of a breach of security;
- security policy – developing a policy that implements measures, practices and procedures to reduce the identified risks to security;
- staff training – training staff and management in security awareness, practices and procedures;
- monitor and review – monitoring compliance with the security policy, periodic assessments of new security risks and the adequacy of existing security measures;
- looking at Australian and international standards as a guide; and
- depending on the size of the organisation and the information it collects, perhaps having an external privacy audit conducted.

Many of these actions are procedural or managerial in nature but impact directly on IT practitioners. The first step in achieving a risk assessment is to identify what information assets are held by an organisation and what threats exist. It is inevitable that online information will need to be identified as sensitive with the aid of IT staff, as well as existing controls on access to the information. Even material such as staff directories, email name and address books, or personal information managers might well contain personal information that needs to be protected.

The risk assessment, once complete, will drive the development of a security policy, plus the development of controls to help implement the policy. There is no single answer to what is a reasonable measure for electronically based systems, though filtering at the boundary between an organisation's network and other networks would have to be regarded as a minimal compliance. Likewise having an audit trail of who has accessed information regarded as sensitive would be judged as a normal reasonable control. A further normal procedure would be defence against social engineering attacks, such as avoiding procedures for resetting passwords without verification of the identity of an individual requesting a change.

If an organisation holds significant amounts of personal information, it is reasonable to expect that a data management plan covering the lifecycle of the data, plus an access matrix, be prepared before gathering the information, and that periodic audits showing compliance with this plan be visible to interested external parties, such as the Privacy Commissioner.

Technological controls are only part of the generation of appropriate security policies. Physical security measures, good training of staff and engagement of staff in

the process are equally necessary as reasonable steps in the protection of sensitive information. Management of an organisation has to make a value judgement about the cost of protection versus the damage caused if protection fails, but they do need to make this judgement in an informed fashion, and it is up to IT staff to contribute in a professional manner in making such decisions.

One aspect of data protection that is often overlooked is the methodology of destruction – both the procedural basis as to when, how, and who can authorise (and other aspects of the Privacy Legislation indicate that data should be destroyed when its primary reason for collection is no longer applicable), as well as the more mundane issues of actually destroying information in a secure fashion – for instance, do you reasonably need to remove such data from archival backups? The answer to that is probably yes, unless there are further controls on access to such archived data.

Another control that most would consider reasonable is the existence of a data spillage plan – what is your organisation going to do in the event of accidental disclosure of sensitive information? The history of incidents in many occupations that have routinely tried to stop leakage of sensitive information indicates that this is the most likely problem to occur. Neither the Privacy Commissioner nor a court is likely to regard imitating a headless chicken as reasonable care.

Review of logs and auditing of security related events are also part of reasonable measures, especially if done by two people. Development of exception reporting strategies and tools is probably more than expected as reasonable at the moment, though as technology and case law accumulates this may be regarded as normal reasonable practice.

In any case, it is worthwhile becoming acquainted with useful standards in reaching an understanding of an organisation's exposure to security and privacy. The high level standards are ISO17799 or AS/NZS 4444 for IT security and AS/NZS 4360 for risk assessment. ACSI33, on the Defence Signals Directorate web site (www.dsd.gov.au) may also give insight, even though the standards espoused there may not be entirely appropriate for commercial enterprises.

One should also understand the use of *certified* products, and the import of certification. The most rigorous international certification standard is known as *Common Criteria* evaluation (see csrc.nist.gov/cc). The purpose of common criteria evaluation is to formally assess the functionality of a security enforcing device against a set of claims (the Security Target or Target of Evaluation) to various levels of trust - rated from EAL0 (least checking) to EAL6 (most checking). It is not worthwhile relying on the rating without also understanding the ST/TOE.

It is not a panacea to make use of certified products, but use of certified products in the manner evaluated will certainly satisfy the reasonable measure claim. You should also note that understanding the methodologies concerned to arrive at a rating requires more than a small amount of effort (and pain).

In this regard, it is also a reasonable measure that review of risk is done on a periodic basis as the threat environment changes, and that IT security practitioners have information sources which alert them to significant threats or countermeasures (such as review of networks, security checklists and tools for OS and application hardening, use of safer architectures, and so on).

There have been some moves to ensure that IT security practitioners are themselves certified in some fashion. Currently, this is probably not reasonable to expect, but one needs to advise, watch this space.

In addition to the specific security requirements under NPP 4.1, the Act also poses security challenges in other areas.

Under NPP 6, individuals have the right to access any information that the organisation holds on them. The manner in which the organisation provides access, whether in digital or hard copy form, is at the discretion of the organisation. The organisation would need to satisfy itself that the person seeking access was who they say they were, which would obviously require some form of verification of identity.

The question of identity verification is far from clear, especially since the guidelines do allow for pseudonymous collection of data – that is data which is personal but not directly identifiable as a particular individual without additional knowledge which might not be held by the organisation which holds the bulk of the personal data.

However, it is quite clear that if an organisation is going to provide electronic access to individuals, it has a clear responsibility once again to take reasonable steps that the data is made known only to the individual concerned. Thus one will need an identifier that is assigned on proof of identity and ideally is only useable for a limited period or particular enquiries. Moreover, it is also clear in the legislation that this identifier must not be the same identifier as is used for other purposes, such as Driver's licence, Medicare Number, Loyalty scheme identifier, or financial account number, to name but a few. Use of Public Key Infrastructure is the most viable solution at the moment

So, what steps should organisations be taking to ensure that they comply with NPP 4.1

First, businesses must ensure that they have a privacy policy. They must examine the Act, the NPPs, the Guidelines to the NPPs that have been issued by the Office of the Federal Privacy Commissioner, and apply these to the nature of the business, and what it does with personal information, and produce a practical privacy policy that suits the business and is within the law.

In addition, most importantly businesses must have a means of verifying the identity of the person who is seeking access to their personal information. How do they ensure that Bernard Hill is the Bernard Hill who has information within the business?

Business must also have a means of retrieving all the personal information that it holds about a person. This will pose particular problems for large organisations that may have several databases, perhaps scattered around the world. Such a business must put in place a means to gather together all of this information and provide it to me upon request.

Finally, a business needs to work out what vehicle that it is going to use to provide me with access. Is it going to allow me to log in to a web site and put in my password and retrieve it that way? Or is it going to require me to come to a front counter, be given a file containing all of my personal information, then taken to a private room to view and photocopy it?

These are most of the practical issues that business will need to resolve on access rights alone, before 21 December this year.

ISO/IEC AS/NZS4444 and E Commerce

Gary Gaskell

1

Contents

- Introduction
- Issues
- Conclusions

2

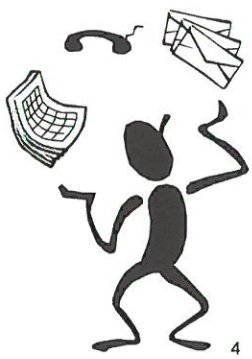
ISO/IEC AS/NZS 17799

- Information Security Management Standard
- Part 1 - 1999
- Part 2 - 2000 (AS/NZS 4444)
- Based BS7799
- BS7799 based on industry - Shell Oil etc

3

ISO/IEC AS/NZS 17799

- Catalogue of controls
- Recommended baselines
- Risk based assessments



4

Best Practice Security Reviews

- Policy that defines the required protection level of assets
- Known exploitable vulnerabilities fixed
- Strong accountability
- Enough redundancy for machines in hostile environments

5

Issues

6

Definitions

- "appropriate"
- Not appropriate for interconnections
- Vulnerability, threat & risk used differently in industry
- Not really a standard - not enough guidance to pass a "repeatability test"

7

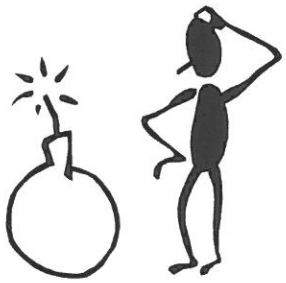
Risk Assessments

- Catalog limitations
- Why not just AS 4360:2000 ?
- Full resort to risk assessment = no saving due to standard

8

Threats

- Internal threats - assumption
- External threats in ecommerce
- Threats:
  - Assets
  - Agents
  - Opportunity/Method



9

Misuse of Standard

- Management standard v's technical specifciation standard
- Compliance of a particular service

10

Effectiveness & Completeness

- Effectiveness != lack of known security breaches - particularly in ecommerce
- No acknowledgement of imperfections in controls

11

Classification

- Only a requirement that a system exists and an "appropriate" access control system exists
- Default classification system could be very useful

12

Evaluation

- Requires new systems under go testing
- Testing cannot prove lack of faults - Dijkstra
- Development process based assurance not required

13

Cryptographic Quality

- "Good" key management is required
- No mention of randomness for key generation
- Real problem - e.g. Netscape in 1996
- WEP in 2001

14

Known Vulnerabilities

- Unpatched systems cause most break-ins - CERT
- 100 advisories from one software vendor in 2000
- Standard does not require known exploitable vulnerabilities to be patched

15

Malicious Content

- AS4444 is usually not specific
- Mentions malicious software
- Malicious data - e.g. buffer overflows
- Buffer overflows - The reason for 50% of breakins over last 10 years - CERT
- Continual catch-up game

16

System Integrity

- Mentions digital signatures are available
- Mentions SSL
- Requires "appropriate" mechanisms
- No mention of tripwire type tools
- Tripwire etc required by CERT, IETF recommendations

17

Electronic Commerce

- Section raises 9 options
- Network access control required
- Compared to BSI, IETF
- Light touch
- No real specification of security offered
- Industry develops own standards

18

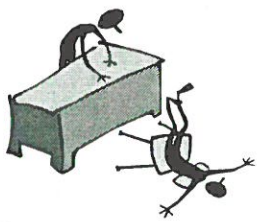
Authentication Quality

- Standard concerned about passwords rather than "authentication quality"
- Passwords - 6 characters are good enough
- No difference in passwords required for privileged users
- Passwords are out dated

19

Logs & Accountability

- Backups are important
- Should mention:
  - Backup full system
  - Back before deployment
  - Backup logs
- Needs more protection of logs
  - See 'rootkit' features



20

Policies

- Very little guidance
- Monolithic policies are too complex to handle
- Policy:
  - Business Security requirements - The objectives
  - Security architecture
  - Operations and detailed technical configuration

21

Miscellaneous

- Advisories services
- AusCERT
- Known construction vulnerabilities

22

Conclusions

- Management framework does not specify the level of security to be provided
- No repeatability
- Catalog of controls is incomplete

23

AUUG Paper: Management of IDS Data

Nathan Carey

Information Security Research Centre, QUT,
Brisbane, Australia
carey@isrc.qut.edu.au

Abstract:

Intrusion Detection Systems (IDS) are a relative newcomer to the security scene, and with the recent interest in IDS, has come a rush of intrusion detection products to fill the niche. Unfortunately, like many recent technologies, the emphasis has been on getting products to market, rather than the usability of those products. This paper aims to look at issues with managing IDS, a comparison of recent advances in the field, and proposes a solution in terms of a generic IDS management framework, MAID, allowing for more sophisticated management, and centralised administration.

1. Introduction

With the prevalence of IDS in the marketplace and research arena, many systems administrators have a wide range of solutions to choose from. For many, the choice will be either one homogenous commercial IDS system or a collection of free tools to do the job. However, while the single commercial system seems attractive, rarely will the system either cover the entire range of systems required, or perform well enough that other measures are not required. This leaves a situation of many implementations containing two or more IDS, complete with their own management frameworks and separate databases for storing alerts. While many of these systems can communicate with Network Management Systems (NMS's) by means of SNMP traps, this support is usually only built into Network Intrusion Detection Systems, and virtually all are commercial. This may not affect many networks, but a growing proportion of setups are choosing to use Open Source or free software because of the ability to compile and run on non-commercial operating systems, or because modifications must be made to support a specific platform.

The upshot of the lack of cohesive cross-platform management structures is the ability to effectively manage large IDS networks. In fact, even some commercial systems' management interfaces do no adequately deal with the amount of alerts generated by large networks. Much like SNMP did for the network management sector, a common framework for the monitoring of intrusion detection, and even certain security devices, is required.

The remainder of the paper discusses the process required to fulfil this need, and a prototype that attempts to do just that. Section 2 discusses background information for IDS, such as useful standards and the reasons for the research. Section 3 discusses the concepts behind IDS management, such as aggregation, correlation and reporting. The paper continues with details of our implementation - Management Architecture for Intrusion Detection (MAID), and finishes with conclusions and future work.

2. Background

Most IDS generate a large amount of alert traffic. Systems such as Snort can generate thousands of false alerts every day, even on moderately loaded networks[1]. This leads to a need for a management framework that can both support diverse alert formats as well as provide a protocol for transfer that provides security and functionality for a broad range of applications. Two current standards in this area from the IETF provide this functionality well, the major concern is that of a management framework using these standards that is free, to promote usage, as well as useful enough to actually be used in place of both proprietary and site-developed systems to deal with this issue. In this section we describe relevant standards that deal with IDS interoperability, and so affect the process of managing IDS.

2.1 CIDF[1]

The Common Intrusion Detection Framework was initiated as a way for certain of the DARPA funded intrusion detection initiatives to communicate effectively. CIDF produced useful work in allowing IDS to communicate, and the requirements that would be needed to allow meaningful exchange of information. This also seems to have been the precursor to work from DARPA and Boeing in intrusion detection, such as the Intrusion Detection and Isolation Protocol (IDIP)[2] and Active Networks Intrusion Detection and Response (AN-IDR). Also, much of the work put into CIDF was used to start the fledgling IETF IDWG.

2.2 IDWG[2]

The IETF's Intrusion Detection Working Group (IDWG) deals with interoperability issues in intrusion detection. This effort was largely constructed out of a previous IDS interoperability experiment, the Common Intrusion Detection Framework (CIDF), but dealt with more issues of standardisation and common formats over the main goal of CIDF which was simple communication. As such, the IDWG worked on two standards, IDMEF which is discussed below, and IAP, or Intrusion Alert Protocol. IAP was based on HTTP-style transfer, and is now replaced by IDXP, below. The IETF work has slowed recently, due to the relative maturity of the work published, but work continues, especially on the issue of including host-based IDS in IDMEF.

2.3 IDMEF

Intrusion Detection Message Exchange Format (IDMEF)[3] is an XML-based solution to the problem of a common standard alert format. The advantage of an XML-based solution is that certain parts can be declared optional, or be extended upon in order to best customise the message to the data provided by the underlying IDS. The IDMEF also allows for configurable or changeable standards by adherence to a DTD that can change with different versions. Also, due to the fact that XML is a plain-text format, its performance is a concern, the text may be hard-coded in order to avoid XML overhead. This means, if required, the overhead of XML-based messaging is comparable with normal human-readable messages.

---

[1] http://www.gidos.org

[2] http://www.ietf.org/html.charters/idwg-charter.html

2.3 IDXP

The Intrusion Detection eXchange Protocol[4] comes in where IDMEF left off. IDMEF does not cover issues of transport or handling of messages, and IDXP is the IDWG's solution to this. Based on Blocks eXtensible eXchange Protocol (BXXP/BEEP) IDXP is actually an XML-specified protocol, and allows for a range of flexibility. IDXP's core duties include the transfer of messages, session establishment and security. Implementations of the protocol are currently in development for Java and C++ to enable libraries for public use.

3. IDS Management

IDS Management as a term itself needs to be clarified here. In this case, we speak of IDS Management being the both the handling of IDS data in a proper way, as well as the ability for a management system to have some modicum of control over the IDS themselves. In this prototype, we deal with the first area, with the possibility to break into the second at a later stage. The issue of management becomes more of an issue when additional features are added to an IDS, such as the ability to act as a central resource for security events, as in IDS/A[5] and when response to intrusions and sharing of intrusion information become more prevalent. Also, a central management structure allows for architectures such as IDIP[2], CITRA[3] and the STAT[6] architecture.

In fact, the STAT architecture provides many of the requirements of a management infrastructure in the form of the CommSTAT and MetaSTAT[7] architectures. The CommSTAT architecture uses IDMEF for communication and SSL for encryption, while the MetaSTAT system uses a central database and control features to manage sensors. While this is useful, the scope is slightly limiting, and does not concentrate on the ability to use common IDS within the framework. While a great achievement, it is our contention that the STAT architecture can be improved both in design and functionality to bring greater IDS independence to the system, as well as improve standardisation and hopefully performance. Also, the primary goal of our work is the collection and analysis of data, both in terms of real-time filtering and correlation, as well as off-line methods for analysis of trends and the ability to track series of actions.

3.1 Alerts and Attacks

Many attacks rely on the ability to overload the system either with erroneous alerts, or hide relevant information within a great amount of bogus attacks.[1] In order to avoid this, we need filtering mechanisms, and the ability to directly control our IDS monitoring real estate. One way to accomplish this is to use two main methods of viewing attack data:

a) Real-time alerts which are filtered down to allow for only the most relevant to be shown, and
b) Off-line methods initiated by the administrator to show trends in alerts, distribution of alerts, and the significant alerts over a period of time.

This is the core functionality of MAID. To provide these features in an efficient package, the architecture must deal with the issues of aggregation, correlation, reporting and analysis, and provide a common framework that is easy to incorporate into current IDS designs.

---

[3] http://www.iaands.org/discex_II/Briefs/12June/D&R/D&R_4_DISCEX-II-CITRA.ppt

3.2 Alert Aggregation

Alert aggregation allows us to converge all our alerts into one central source with no loss of information. The easiest and fairest way to accomplish this is to transfer the alerts of all IDS into a common format, which encompasses them all, and store this in a database. Also, by aggregating information, we can secure both it's location and storage more easily, which is important for forensic issues, as well as correlation, below. Alert aggregation's main concerns for implementation are:

a) Communication to a central point
b) Maintaining secure communication
c) Scalability
d) Routing of Alerts

The only new concept here is the routing of alerts, which is the ability to place alerts in whatever portion of data storage is required. This can mean that one central server can handle and segregate information gathered from many different sources, based on rules setup if required by the administrator. The functionality can be required by some organisations who have differing security requirements for alerts, but do not wish for different systems for each different security level.

Of these concerns about implementation, IDXP can handle a) and b), scalability is handled by careful design, and routing of alerts can be performed by limited interrogation of alerts to determine where alerts should be routed.

3.3 Alert Correlation

The correlation of alerts is used to identify relationships between either alerts, sequences of alerts, or the data contained within the alert. In a management framework, it enables complex analysis on alerts to be performed, without burdening the IDS itself.

Correlation is normally performed at two stages – real-time, or when the alert is generated, and off-line, sometime after the alert is generated. The real-time system suffers from problems of scalability and the time window used for correlation due to the performance concerns with real-time analysis. Off-line methods do not have the same performance requirements, and such can support a far greater range of analysis and timeframes.

In our model, the correlation of alerts is performed in both stages:

a) Real-time Correlation

   The problems of real-time correlation are generally related to the methods used to analyse data. In our model, real-time correlation will be performed either by STATL[8] or regex-based expressions with a limited vocabulary. It will rely on interrogating certain features present in IDMEF messages, such as source, target and attack, with options for including confidence level (to be included in the next version of IDMEF) and other non-standard features. The main benefit of the real-time correlation can be to filter out a great deal of erroneous data, and to provide the administrator with a way to directly access certain types of alert, rather than having them simply stored in the database.

b) Off-line Correlation

   This is normally provided by analysis either of log files generated by IDS, or the underlying database storing alerts. Neither of these is suited to real-time analysis, due to the large amount of alerts stored in even moderately loaded networks, and the time taken to analyse each item in turn. Especially in large networks, the time taken to do even optimised queries on large datasets can be measured in seconds, which is not appropriate for real-time use when on these

same systems, the average alert frequency is in hundreds of alerts a second. However, both of these methods allow sophisticated correlation to be performed, which could be used to determine sequences of attacks in special circumstances, or post-intrusion forensic analysis.

3.4 Alert Reporting

In properly reporting alerts, the user needs to obtain information in a timely manner, without interruption, and maintain integrity under large amounts of alerts.

We can use a number of mechanisms to give the user alerts:

a) *Pop-up windows / Audible Alerts*
This is the most problematic method, as the use of this can easily become a denial-of service for the administrator. These alerts can be given on high-profile attacks to be given immediate attention. By necessity, these alerts need to only happen occasionally, otherwise the service of closing windows or responding to direct alerts can overwhelm the user. This approach is used on some of the more feature-oriented IDS, mainly for use in low-traffic networks. This sort of scheme is also used for pager, email and windows messaging alerts.

b) *Lists of Alerts*
This method allows the user to see, normally sequenced by time, the alerts that are produced on the system. Unfortunately, these type of systems can be overloaded easily, and do not give a great deal of information, especially if used on all alerts where low-profile alerts can dwarf the more important issues. This sort of approach has been popularised by simple management systems, because of the ease of translation from average log files.

c) *Formatted Tables*
Formatted tables can be used to display information in a more content-based approach than lists of alerts. Normally tabular approaches are used in HTML-based interfaces, and contain information in a structured manner for the user. Structuring could occur on the attack listed, IP address, priority etc. This is also used by systems whose store their data in databases.

d) *Tree Structures*
In more advanced systems, tree structures are used as an alternative to tabular structures. This provides a more intuitive approach, and due to the normal tree features of hidden branches, also allows for a great deal of information to be displayed in a small amount of space. Unfortunately, usually these systems take up a lot of screen real-estate, and are not built with scalability in mind. However, they remain a display mechanism that is gaining in popularity and should supersede tabular approaches in many sophisticated interfaces. Some, such as MetaSTAT, offer this as an alternative interface to a table. This is made easier by the Java language used for the GUI.

e) *Graphs*
Graphs allow for a sophisticated way to analyse trends and series of data in a quick manner. While not generally useful in an individual alert sense, many large institutions use interfaces such as these to analyse large data sets to determine relationships between data, and to help visualise the state of the network.

f) Statistics

Statistics are used very often in tabular interfaces, but can be used on their own to good effect. Statistics, much like graphs, can help in analysing trends, but careful design can also allow for individual alert analysis. However, the non-intuitive nature of analysing large amounts of pure numbers, can be detrimental to many users.

3.5 Alert Analysis

Analysis of alerts deals with classification, tabulation, and evaluation of key information and features in order to ascertain meaning and probable effect. While this is an oft-used term, its application to IDS is generally under-utilised. In general terms, we can use alert analysis to determine trends, distributions and divisions in our alert data. This is generally displayed in tabular, statistical or graphical form to be of most use to the administrator.

Analysis of alerts is a complex issue that is still relatively new. Until recently, the acceptable form of analysis was a total of how many alerts had occurred, and storage of flat lists of alerts for the user to analyse, sometimes by hand. More recent advances have lead to the storage of alerts in databases, allowing for complex analysis to be performed in a relatively timely manner. Unfortunately, the science of how to properly analyse these logs is still new, and few full-featured analysis tools are available. The construction of a portable set of analysis structures has long been hampered by the fact that each IDS logs a different set of information together with each alerts, sometimes in completely different formats. IDMEF can help to solve this issue by standardising the information contained in an alert, which can help to standardise the way it is analysed, and so improve the field.

4. Real-time Alert Pre-Processing

4.1 Filtering

In filtering information, we need an expressive language capable of adapting to a wide range of features available in the average alert message. If using IDMEF, we have the ability to look at XML either as an XML document, or as a plain text document, and match information based on that. A simple example could be *"delete the alert if it matches alert "PortScan" from source "192.168.1.1"*

This can easily be abstracted to the familiar *"perform action(s) X based on Y content"* and can be easily gathered from the XML message. The simplest form of this could be to match actions with regular expressions, which could be generated either by an automated process based on criteria, or programmed by the user themselves. It is this approach that is used for the rudimentary level of pre-process filtering of messages. The basic functions supported are *delete*, *forward(a,b..)*, and *escalate(i)*. Delete is self-explanatory, forward is used for simple routing to other processors, and escalate is used to increase the priority of certain alerts by a pre-determined amount, if required. This can also be used to increase the alerting capability of certain types of alerts when under attack. Any combination of these can be used, with their processing being performed in reverse order, i.e. A function labelled with *delete, forward, escalate* would first escalate the priority, forward it to another processor and then delete it from it's processing queue.

4.2 Rule-matching

One of the advantages of a cohesive management framework is the ability to process alerts, as a sort of holistic IDS. By this we mean the matching of sequences or patterns of alerts that might indicate large-scale attacks, successful penetration of a system, or something as mundane as the spread of a virus. By the addition of rule-matching on alerts themselves, we can perform this sort of processing and improve the ability to monitor large-scale IDS implementations. The best mechanism for specifying generic rules at this point in time is UCSB's STATL[8]. STATL allows for transitions between state to be described, and a sequence of transitions linked to provide a signature for an attack. The general nature of STATL allows for it to be easily adapted to alert information, and a scenario editor exists to provide an easy way to describe scenarios.

Work by Hervé Debar and Andreas Wespi [9] also includes a specification for an Aggregation and Correlation architecture, concepts of which will be included in due course.

4.3 Routing

Routing of information becomes an issue of great concern in large networks – who do we give data to, and how many copies of it are stored? The MAID architecture includes the ability to route information both on priority, source and target to enable differing views of IDS alerts to be represented on multiple stations. This functionality can be incorporated into both the IDS sensor and the processing station depending on the separation of data required. The most efficient method would be to implement this feature on the processing station of the IDS, and have it route information to other stations, but this may not fit the security requirements of some networks, so an IDS should be able to be configured to send information to multiple destinations. The usage of routing on both could be used for certain divisions to implement duplication of alerts that are invisible to other stations. If correlation is occurring between these stations however, either the algorithm needs to be intelligent enough to recognise and ignore duplicate alerts, perhaps by alert ID, or a registry of the transport of messages should be kept and referred to in case of duplicate messages.

5. Our proposal - MAID

MAID is designed to provide a cohesive management framework based on commercial and research IDS. In order to support MAID, many systems will require modifications to translate alerts into IDMEF format, and potentially modifications to allow for translation of MAID control messages into native reconfiguration commands, if desired. MAID is a simple multi-level hierarchy allowing routable delivery of messages through management chains and easy storage through a relational database. One of the aims in developing MAID was to use as many existing standards as possible, both to improve the quality of the final product, as well as increasing modularity and pluggability with other systems.

5.1 'Cognisant' & 'Passive' IDS

One of the core aspects of the system is the idea of 'cognisant' and 'passive' IDS. Cognisant IDS are those with knowledge of the architecture, and the ability to interact meaningfully with the system itself – both in terms of active messaging, as

well as additional features. The average cognisant IDS will register the features it supports with the architecture.

Passive IDS are those with no knowledge of the architecture. They will generally need an alert-IDMEF translator as well as a translator for additional features if available. This will be accomplished by the idea of an IDS Interface proxy, described below.

5.2 Requirements

a) Produce a framework for allowing multiple IDS to communicate to a central console. The prototype utilising the framework will perform the following tasks:

i) Utilise common standards for communication
ii) Use common standards for representing alerts
iii) Provide secure, authenticated communication over insecure channels.
iv) Increase the current functionality of major IDS such as Snort and Dragon.

b) Produce a processing engine for the information in a) that will allow for intelligent analysis of IDS data to produce the following results:

i) Correlation of attacks over multiple systems
ii) Reduce the amount of repeated or duplicated alerts
iii) Classify levels of alert
iv) Organise information into logical hierarchies to allow for easier monitoring
v) An approach utilising as many standardised features as possible.

c) Produce back-end processing methods that will allow post-processing of data. This will include:

i) Database with query language support
ii) Normal storage such as text file available

5.3 Design

A diagram of the architecture is included below.
The major concepts are the Management Interface, the Processing Engine, the Upstream/Downstream Normalisers and the IDS interface proxy.

a) *Management Interface*
The management interface in functionality will be much similar to Stage 1, but will have control for real-time alerts via the transaction agent, below.

b) *Processing Engine (PE)*
The PE takes in information from the various IDS sources, and sends them to the Main Data Store. This enables the transaction agent to handle connection establishment and any extra functionality required above what a normal database provides. Also, perhaps more importantly, the transaction agent can intercept messages on the fly and determine if they should be given directly to the management interface as well as the data store. This can enable real-time alerts to be given to the user.

c) *Upstream/Downstream Normalisers & Comm Agents*
The upstream normaliser is performed by the combination of translation into an IDMEF-formatted alert by either the IDS or the IDS Interface Proxy. It's role is to transfer alerts to the PE either in batches or real-time, depending on application. The downstream normaliser is used for the transfer of control information, which is still in development. An initial version will use the STAT control signals as a proof-of-concept system, but work will include the

specification of more general controls, perhaps suitable for other systems aside from IDS. This will enable the system to perform response capabilities.
Comm Agents are those given to the Cognisant IDS. Their task is to simply send messages to the transaction agent. Ideally, in an IDS with knowledge of the architecture this could be incorporated directly. Each comm agent should register the IDS with the Management Interface in order to properly deal with its information. The communication system is simple at present, but will incorporate IDXP when standardised.

d) *IDS Interface Proxy & Interpreter*
This is the core of integration of Passive IDS. Essentially, the IDS interface needs to have knowledge of the workings of the IDS plugged into it, as well as how to translate the messages into the common format, and how to prioritise messages. It will require plug-in modules for each type of IDS, and each IDS will need to be registered, in order that the proper translation mechanisms be performed. This registration will then be passed onto the Management Interface for the user.

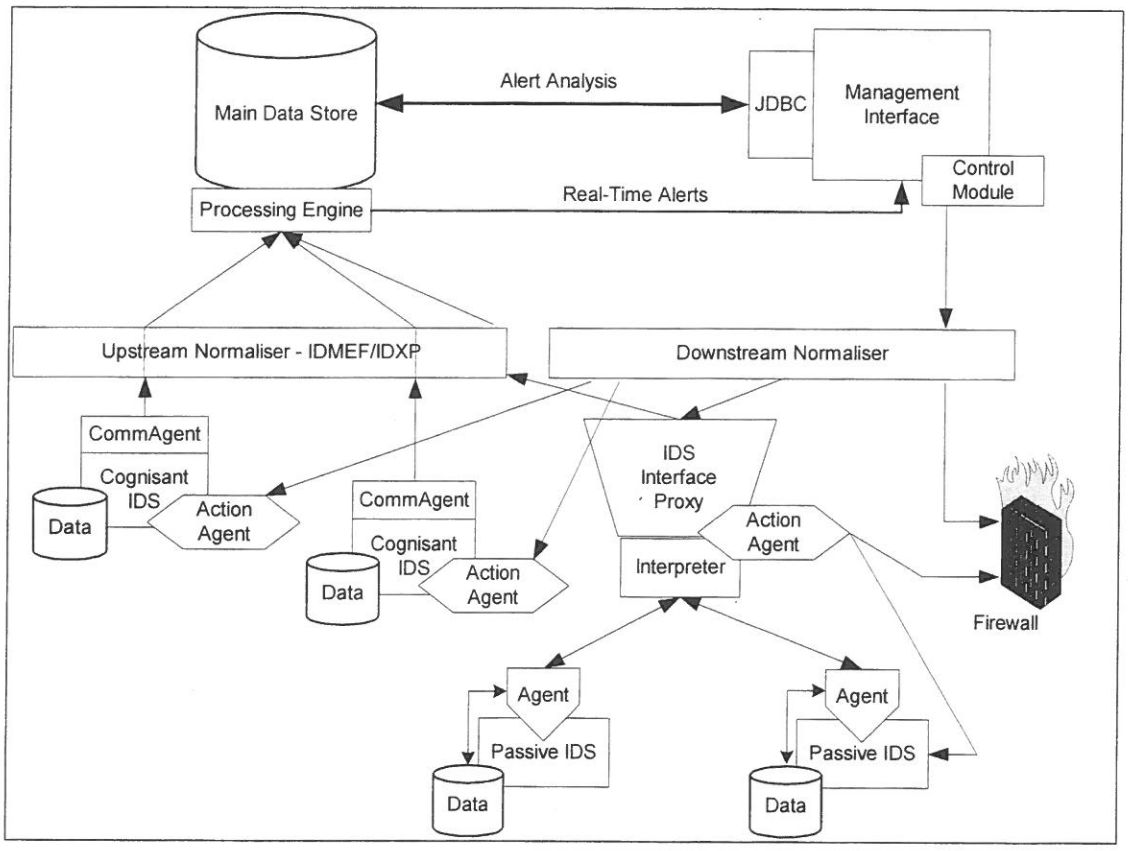


Figure 1: MAID Architecture

6. Software Implementation

6.1 System Profile

The system has been implemented with alert aggregation, database-driven correlation, and a management interface to control the system. Java was chosen as the programming language, both for the general speed of development, as well as platform independence and handling of XML data. The current systems used are

Snort and Dragon. Host-based IDS will be integrated after the next round of modifications to better support host-based systems that have passed IDWG vetting and are standardised enough to implement. The real-time aspect of the processing engine is rudimentary at this stage, with simple regular expressions being used, with the intended capability to move to STATL representation in the near future.

6.2 Software

As mentioned above, Snort and Dragon are the first IDS to be implemented, for the purposes of adding a management structure to primarily CLI-based alerting systems. The XML data transfer is accomplished by using the Xerces[4] library from the Apache project, and IDXP experimentation is being performed using the Java-IDMEF library[5]available from Sourceforge in beta stage. XML to relational database transfer is using the XMLDBMS[6] libraries from Ronald Bourret and a basic mapping file from the STAT project. All database access is accomplished using the JDBC interface, and displayed using table and tree structures on a Java GUI. The database itself is Postgres[7] while MySQL could be used, the ability to constrain access to the database based on predefined views gives the application the ability to provide more complex and easily administrable security than available on MySQL. While this selection of software may not provide the best performance, it does suit the aims of the project admirably, in terms of portability, usability and application development. It will be released under the GNU GPL[8].

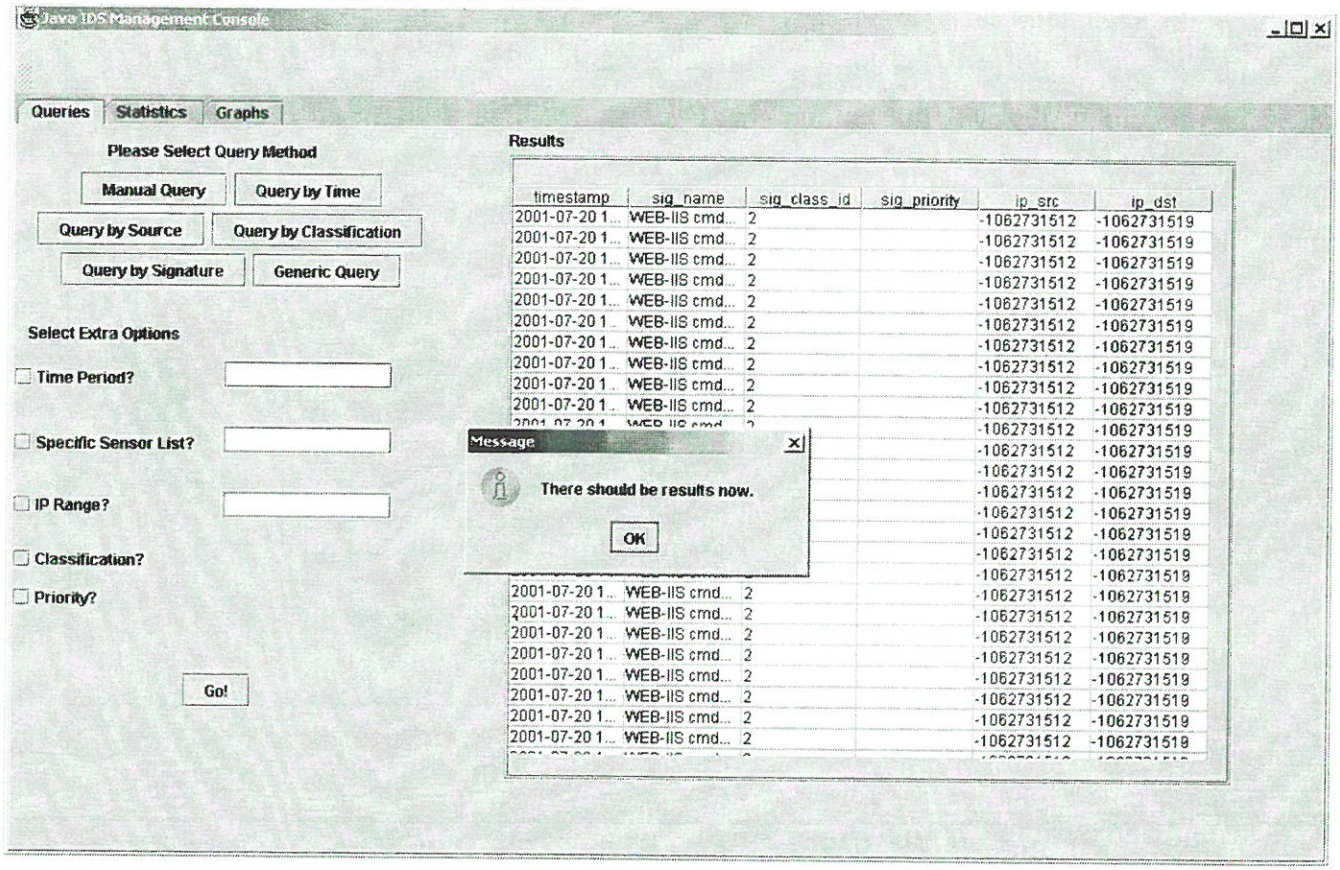


Figure 2: MAID Operation

[4] http://xml.apache.org/xerces-j/index.html
[5] http://sourceforge.net/projects/idxp-java/
[6] http://www.rpbourret.com/xmldbms/
[7] http://www.postgresql.org
[8] http://www.gnu.org/copyleft/gpl.html

7. Conclusions & Future Work

The management interface here is still in early stages of development, but the initial signs indicate that the usefulness of such an architecture, especially in terms of integrating free and research software which is generally open source and poorly managed. For commercial software, the real use comes in the complex management of alert data possible in this system, without the need for a very complicated infrastructure to be in place. The depth of correlation and routing is also very rare, especially for a free system which will deal with a wide range of IDS.

The major areas to be developed are in the pre and post processing of data for correlation and aggregation, especially towards forensic and post-intrusion concerns. Also, the addition of host-based systems will allow the specification of more useful rules which can detect the progress of an intrusion through hosts themselves. The real work now is in the specification and research into the most effective ways to trace intrusions.

The addition of control mechanisms which are rudimentary at this stage is of particular interest, especially in the area of survivable systems and information warfare. While not the core concern of this work, it is conceivable that this sort of functionality could be added at a later stage.

Overall, the architecture promises to be useful and useable in many network IDS deployments, but much work needs to be done before the system can be classed as 'enterprise-ready' which is the eventual goal.

8. References

[1] Hoagland, J.A. and. S. Staniford. *Viewing IDS Alerts: Lessons from SnortSnarf.* in *DARPA Information Survivability Conference and Exposition Proceedings (DISCEX)* ,. 2000.

[2] Schnackenberg, D., K. Djahandari, and D. Sterne, *Infrastructure for Intrusion Detection and Response*, in *Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX), 2000.* 2000, IEEE: Hilton Head, SC. p. 3-11.

[3] D. Curry, H.D., *Intrusion Detection Message Exchange Format Data Model and Extensible Markup Language (XML) Document Type Definition.* 2001: IETF.

[4] Feinstein, B.S., G.A. Matthews, and J.C.C. White, *Intrusion Detection Exchange Protocol (IDXP).* 2001.

[5] Welz, M. and A. Hutchison. *Interfacing Trusted Applications with Intrusion Detection Systems.* in *Recent Advances in Intrusion Detection (RAID 2001).* 2001. Davis, CA: Springer.

[6] Vigna, G., S.T. Eckmann, and R.A. Kemmerer. *The STAT Tool Suite.* in *DISCEX 2000.* 2000. Hilton Head: IEEE Press.

[7] Vigna, G., R.A. Kemmerer, and P. Blix. *Designing a Web of Highly-Configurable Intrusion Detection Sensors.* in *Recent Advances in Intrusion Detection (RAID 2001).* 2001. Davis, CA: Springer.

[8] Eckmann, S.T., G. Vigna, and R.A. Kemmerer, *STATL Syntax and Symantics.* 2000, Computer Science Dep., University of California Santa Barbara.

[9] Debar, H. and A. Wespi. *Aggregation and Correlation of Intrusion-Detection Alerts.* in *Recent Advances in Intrusion Detection (RAID 2001).* 2001. Davis, CA: Springer.

Commonwealth Bank

Secure banking

Dr Steve Anderssen
Executive Manager, IT & T Security
Group Technology

November 2001

- **Banks and security**
- **What is security?**
- **Security management**
- **A target security infrastructure**
- **"Securing the Environment" today**
- **The key challenges**

2

- **CBA web site is an easy hack**
  19 July 2001 Illawarra Mercury
- **QuickLine banking hack can only bite the careless**
  24 July 2001 Australian Financial Review
- **Nimda computer virus hits NAB**
  20 Sept 2001 Australian Financial Review
- **Westpac site collapses as users try to log back on**
  17 Aug 2001 Australian Financial Review
- **Security fear still a bar to Net banking**
  16 Jan 2001 Sydney Morning Herald

3

Mandatory Business Requirements

- Market of one
- Service anywhere, anytime, anyhow
- Self service
- Absolutely secure
- Single customer ID
- Knowledge driven
- No paper
- Immediate information availability

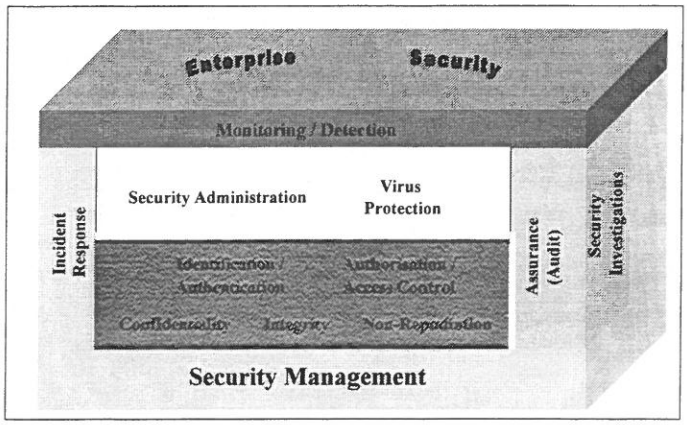
Security is a key business requirement and an important feature of any technology solution for the effective delivery of services

Role of Technology To Make This Happen

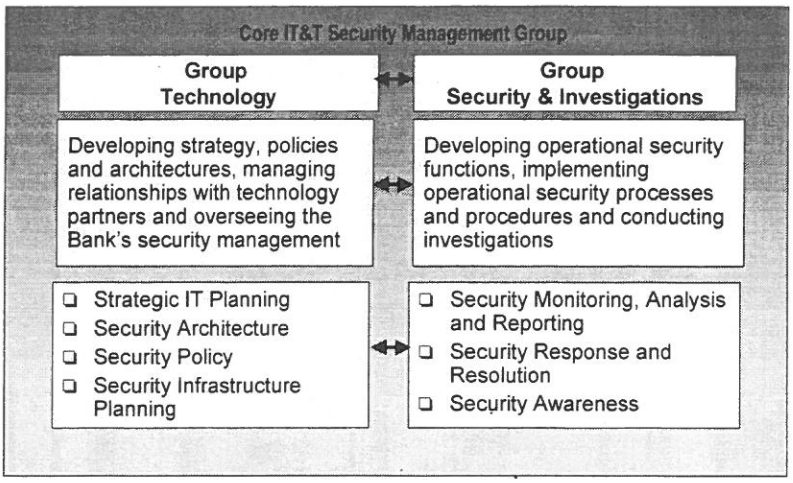
- Browser based
- One integrated network
- Electronic touchpoint
- Best practice security
- Workflow enabled
- Intelligent call centers
- Reusable and scaleable technology
- Common technology & business infrastructure
- 24x7
- Open system standards
- Master directory of data items
- Real-time processing

4

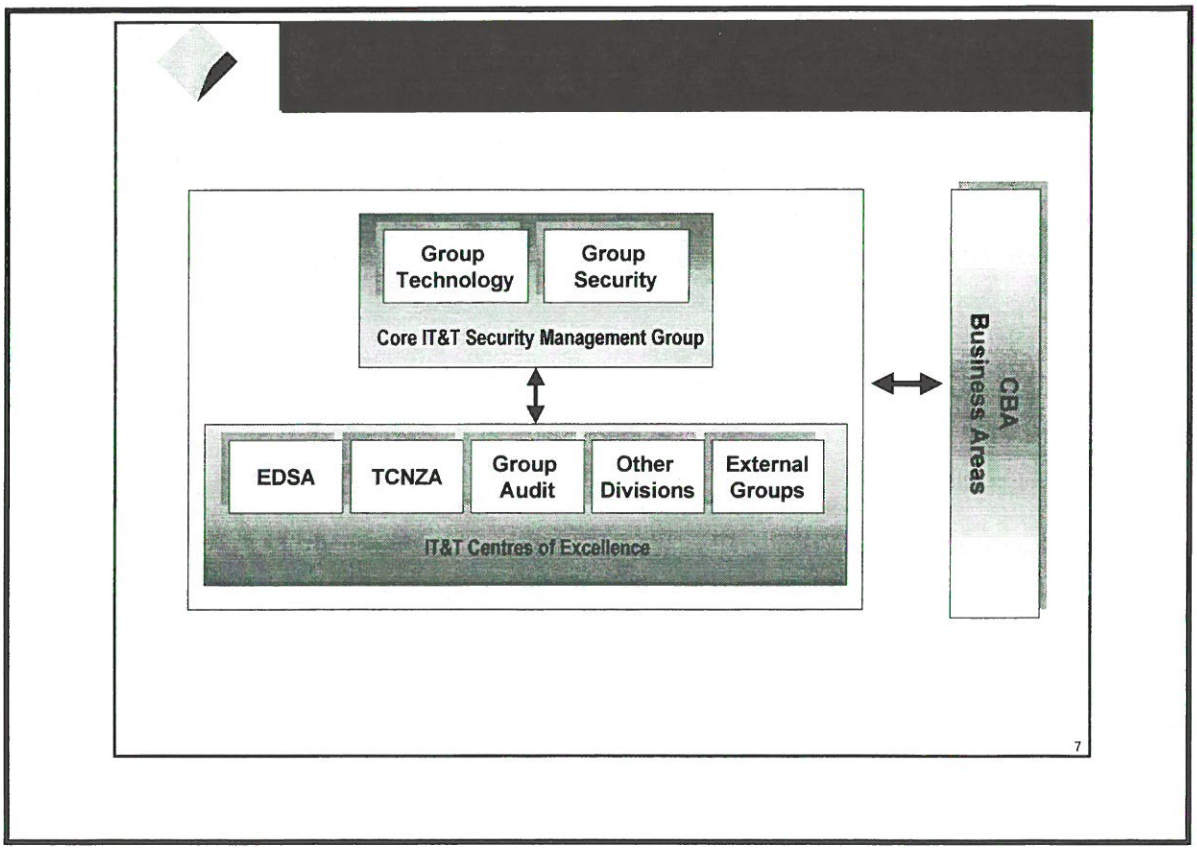
The Security Framework has been used to ensure that all aspects of security are addressed in the delivery of services

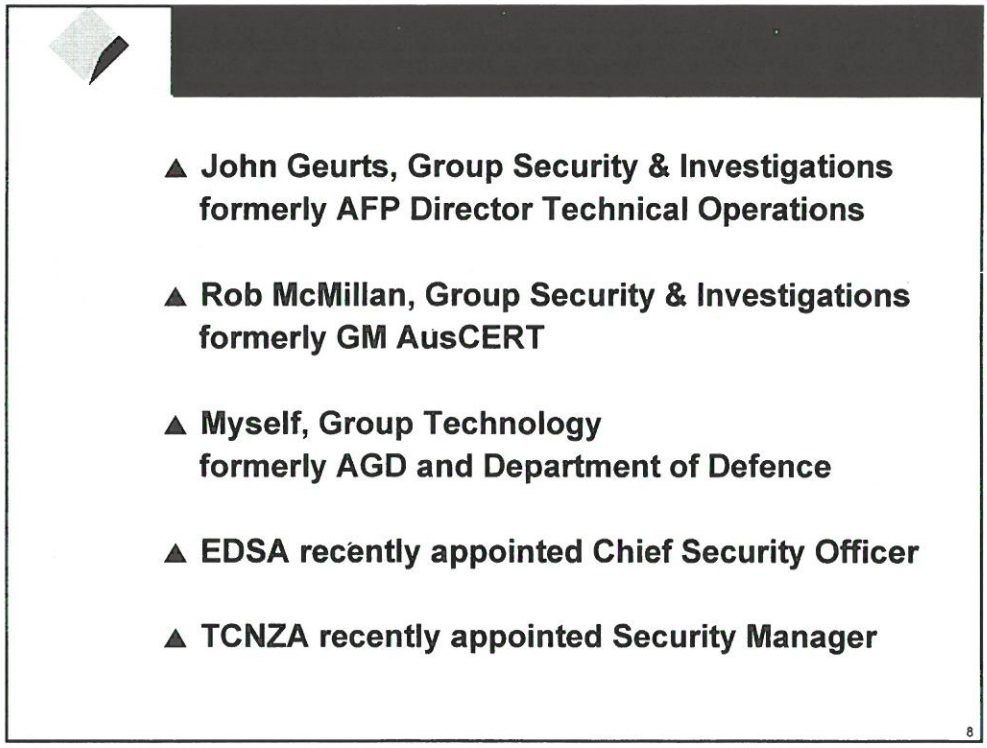


5



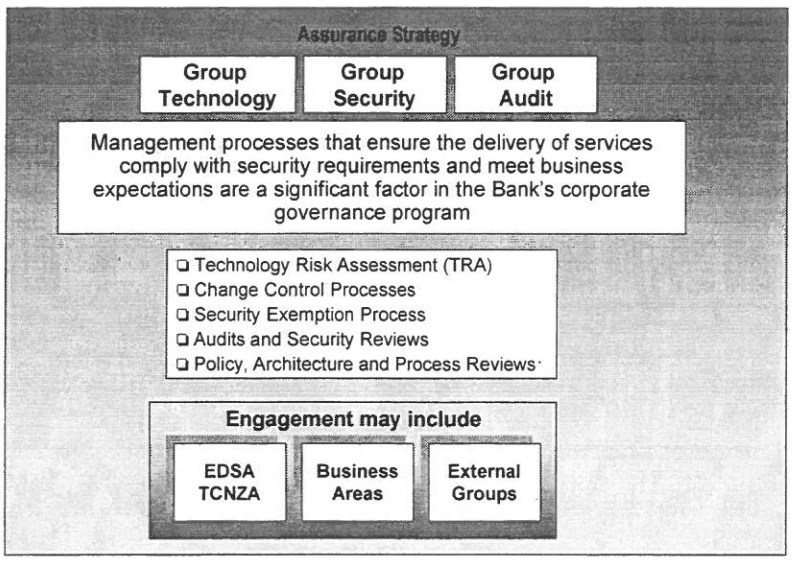
6


Group Technology
Group Security
Core IT&T Security Management Group
EDSA
TCNZA
Group Audit
Other Divisions
External Groups
IT&T Centres of Excellence
CBA Business Areas
7

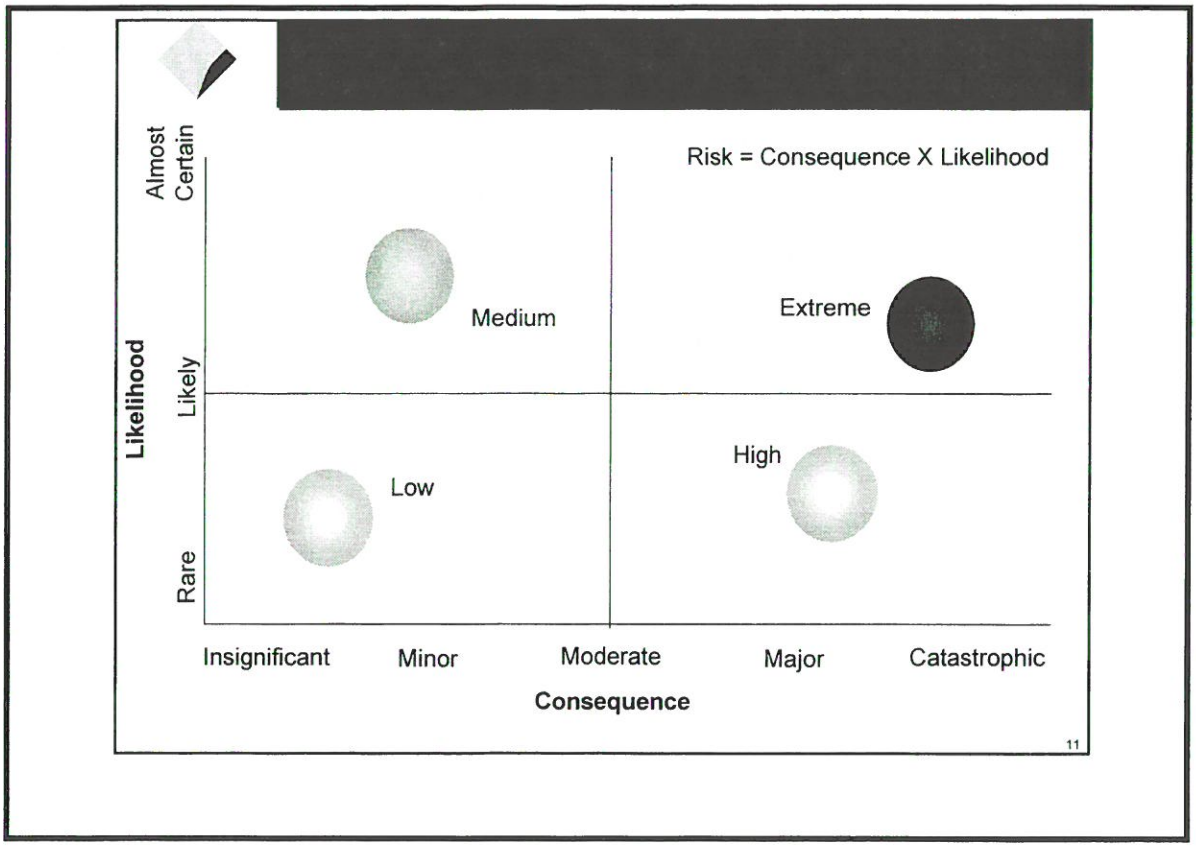

▲ John Geurts, Group Security & Investigations formerly AFP Director Technical Operations
▲ Rob McMillan, Group Security & Investigations formerly GM AusCERT
▲ Myself, Group Technology formerly AGD and Department of Defence
▲ EDSA recently appointed Chief Security Officer
▲ TCNZA recently appointed Security Manager
8

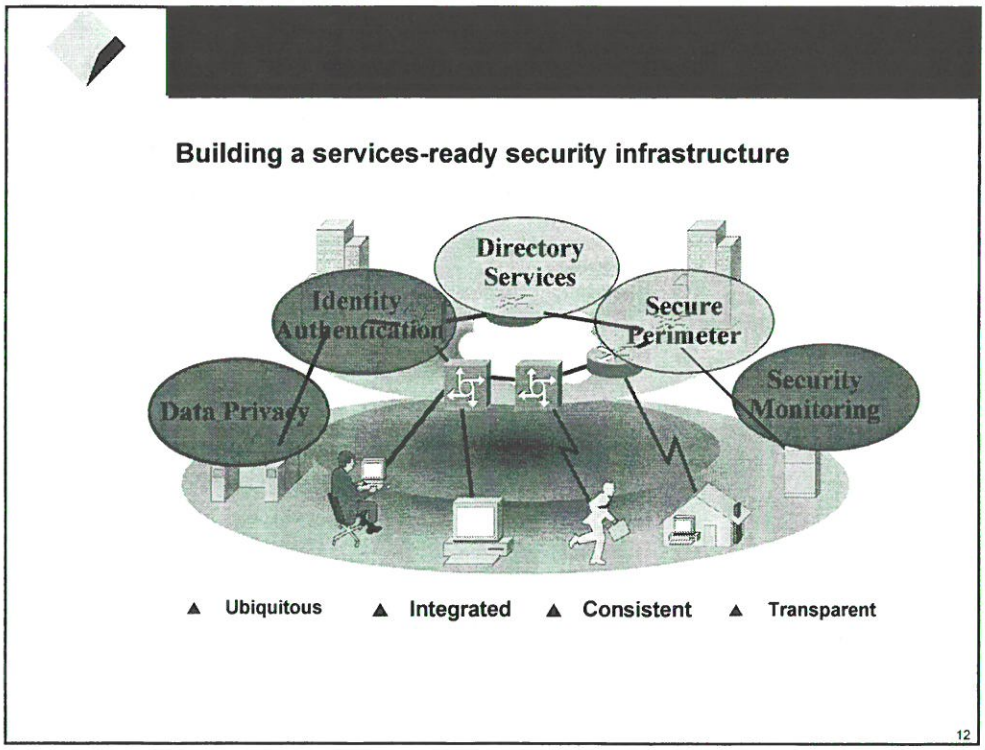
- CBA I & IT Security Policy and Handbook
- CBA Target Enterprise Security Architecture
- CBA Audit Plan
- CBA Incident Response Plan
- Contractual requirements on EDSA and TCNZA
- EDS and TCNZ policies and procedures

9

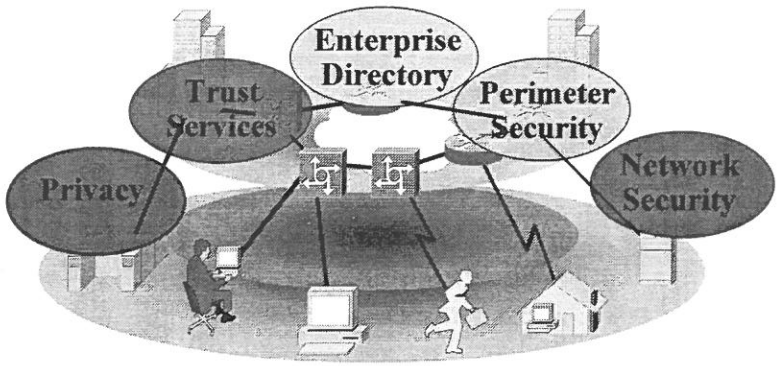


10


Risk = Consequence X Likelihood
Almost Certain
Likely
Rare
Likelihood
Medium
Extreme
Low
High
Insignificant
Minor
Moderate
Major
Catastrophic
Consequence
11


Building a services-ready security infrastructure
Directory Services
Identity Authentication
Secure Perimeter
Data Privacy
Security Monitoring
▲ Ubiquitous
▲ Integrated
▲ Consistent
▲ Transparent
12

The series of projects that will deliver the capabilities of the Target Security Infrastructure



- ▲ Compliance
- ▲ Comprehensive
- ▲ Identrus PKI
- ▲ SSO
- ▲ Secure email
- ▲ File signing
- ▲ Directory Services
- ▲ Dual F/W DMZ
- ▲ VPN / RAS
- ▲ Network IDS
- ▲ AV
- ▲ Host-based logging
- ▲ VLAN's

13

A comprehensive end-to-end review of the Bank's electronic environment

and a program of work to address any issues identified

- ▲ **Divided into thirteen work streams**
- ▲ **Initially only EDSA**
- ▲ **Later will include TCNZA and CBA where relevant**

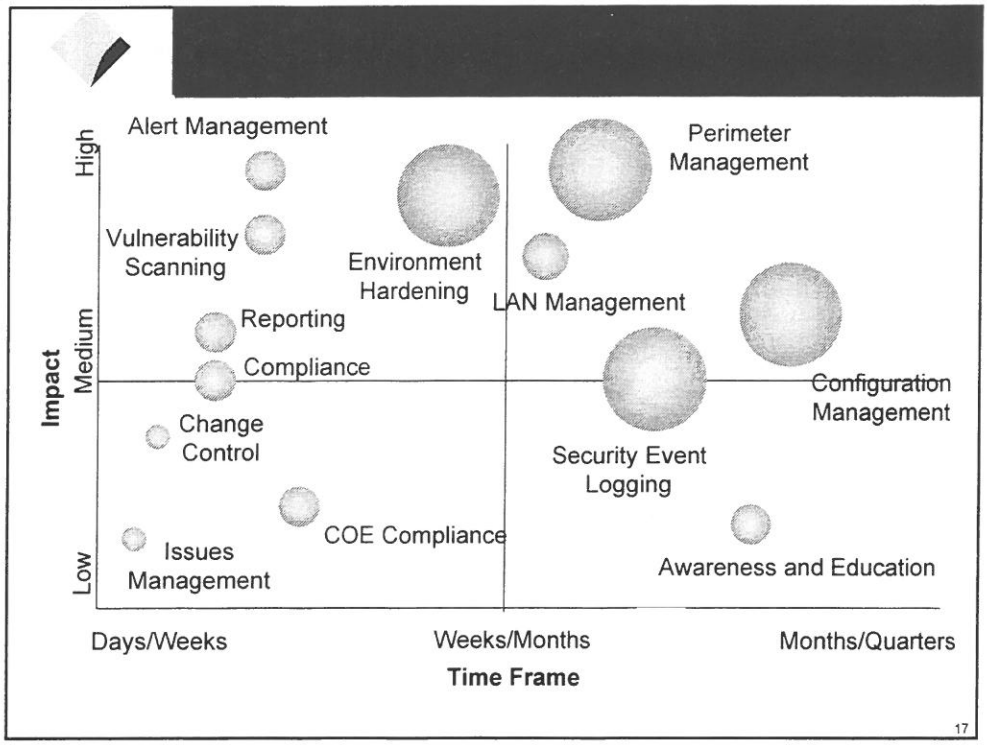
14

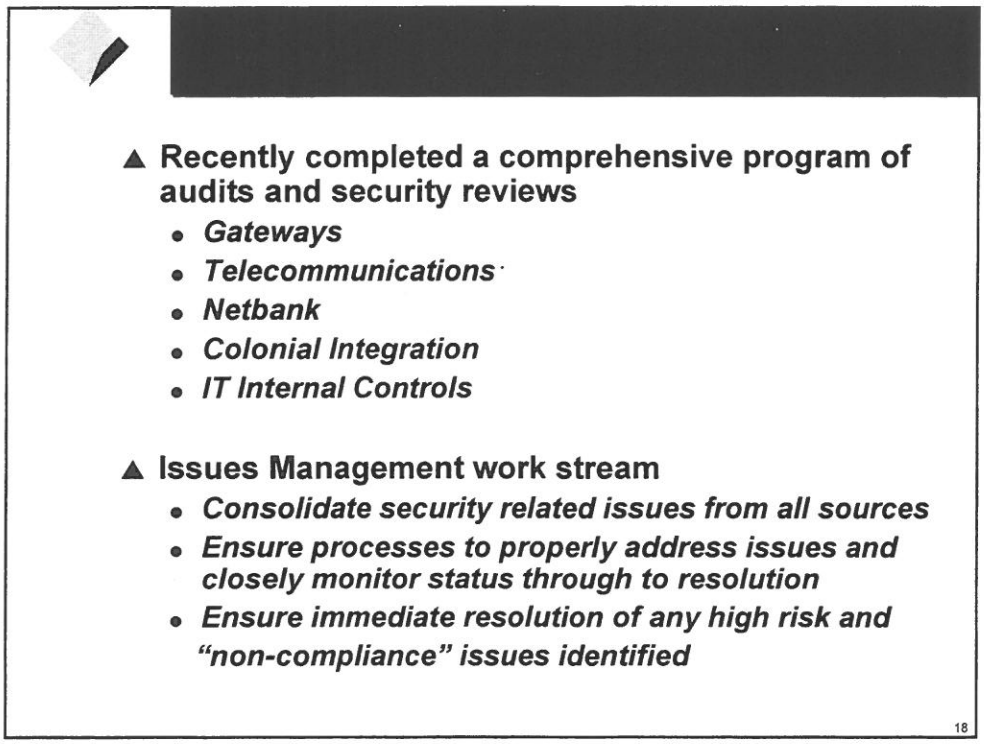
- All services, tools, resources, people and processes impacting on the provision of a secure environment
- All CBA Business areas and all EDS delivery organisations including subcontractors
- All services within base contract and any additional services as required
- Both real and perceived issues
- Preventative and detective measures
- One-time fixes for simple issues and long term solutions for systemic problems

15

- Change control
- Compliance
- Issues management
- Perimeter upgrade
- Security event logging
- LAN management
- COE compliance
- Alert management
- Vulnerability scanning
- Configuration management
- Environment hardening
- Communication and reporting
- Awareness and education

16


Alert Management
Perimeter Management
High
Vulnerability Scanning
Environment Hardening
LAN Management
Reporting
Impact
Medium
Compliance
Configuration Management
Change Control
Security Event Logging
COE Compliance
Low
Issues Management
Awareness and Education
Days/Weeks
Weeks/Months
Months/Quarters
Time Frame
17


▲ Recently completed a comprehensive program of audits and security reviews
• Gateways
• Telecommunications
• Netbank
• Colonial Integration
• IT Internal Controls
▲ Issues Management work stream
• Consolidate security related issues from all sources
• Ensure processes to properly address issues and closely monitor status through to resolution
• Ensure immediate resolution of any high risk and "non-compliance" issues identified
18

- Virus Protection
- Password Management
- Active Content Management
- Wireless Security
- More sophisticated electronic fraud
- Compliance with the Privacy Act
- Customer Awareness and Education

19

- Ph: 02 9378 2338
- Mob: 0438 837 132
- email: steve.anderssen@cba.com.au

20

Wireless Insecurity

802.11b security issues

Neal Wise
nwise@esec.com.au
eSec Limited

#include <stddisclaimer.h>

- I am not a expert on 802.11x
- I do not advocate exploitation of the risks of 802.11x
- I am not associated with any vendor of 802.11x technologies

What is 802.11b?

- 802.11b - 2/5/11Mb
  - Shares spectrum with other technology
- 802.11a - 54/100Mb
  - Devices only recently available
- 802.11g - ? - jury still out at IEEE
  - Unknown fate

Why use wireless?

- Convenience - MP3s in the back yard
- Inexpensive Point-to-point
  - 30km+ is possible given equipment and conditions
- Heritage buildings/wiring avoidance

All of this and more... mostly at the expense of security!

802.11b Devices

- PCMCIA cards
  - ISA PCMCIA/PCI CardBus adapters
- Built-in on some notebooks and handhelds
- Access Points -Routing or bridging devices connecting wireless networks to wired networks

802.11b Vendors

- Lucent/Orinoco/WaveLAN (Apple Airport, IBM)
- Cisco/Aironet Inc
- Symbol
- Linksys, Enterasys, Compaq, Nokia,etc

Lots of vendors are making lots of products

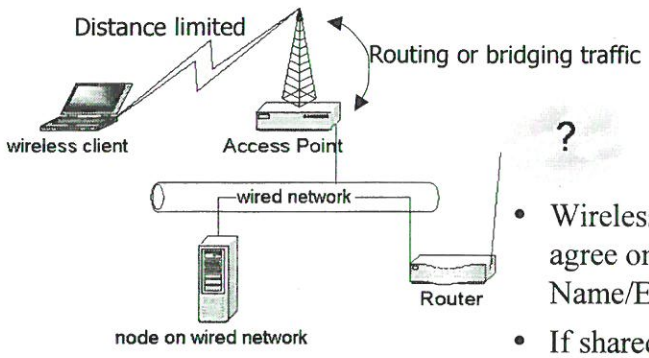
Warning: extremely oversimplified

How 802.11b works

- AP sends out presence information approximately 10 times a second
- 802.11b device appears to be an "ethernet-like" interface on client system - has MAC address, etc
- Some logical agreement in the form of configuration of the client device (laptop, etc) and the AP must occur.

Warning: extremely oversimplified

How 802.11b works - 2



?

- Wireless client and AP must agree on Network Name/ESSID
- If shared secrets or encryption are enabled both ends must agree

How is security considered?

- Link layer
  - Nature of 802.11b requires devices to announce service - disclosing service
  - Agreement of configuration details
    - Default configurations are **often** found
- Network layer
  - Whatever the end user implements - expect the worst... nothing discovered to date indicates security awareness!

802.11b link security

- Early devices implemented no link-level security
- Some proprietary shared-secret technology was introduced. Interoperating between vendors was difficult
- WEP -Wired Equivalent Protocol

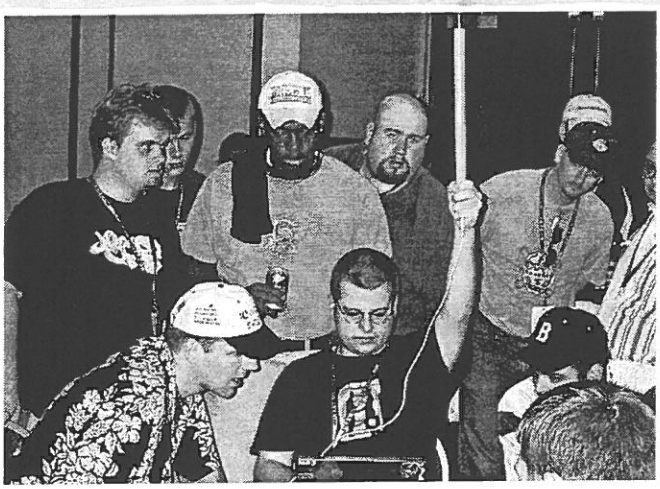
WEP-Wired Equivalent Protocol

- 128/40bit hardware encyrption
- Serious flaws - can be brute forced.
  - "Weaknesses in the key scheduling algorithm of RC4" - paper revealing weakness by Fluhrer, Mantin and Shamir.
  - http://www.cs.rice.edu/~astubble/wep/ -effort by Avi Rubin, Adam Stubblefield and John Loannidis to brute force WEP using findings

WEP-Wired Equivalent Protocol -2

- Doesn't prevent link-level enumeration
  - ESSID, AP name, WEP enabled/disabled still possible.
- WEPPlus. Only Orinoco/Lucent
  - http://www.80211-planet.com/news/article/0,4000,1481_921431,00.html
- 802.1x - IEEE security standard - port based network access control

Wireless Risks - the scary stuff



Wardriving - Wireless Enumeration

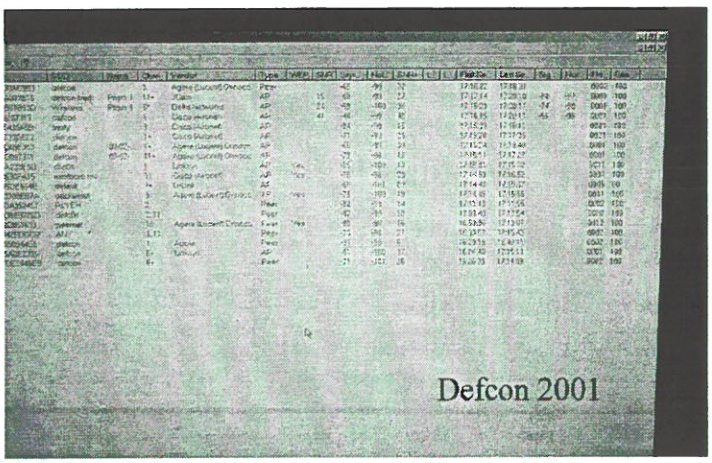
802.11b service discovery

- Netstumbler - for win32 platforms
- "EvilPete" scripts - perl scripts from dis.org
- Dstumbler - for *BSDs (yay!)
- Tools provided with wireless cards!

Wireless Enumeration - NetStumbler

- http://www.netstumbler.com
  - BAWUG/personaltelco.net
- Part of the 'free wireless'MAN movement
- Only supports Hermes chipset cards
- Evil platform requirement... however it works well

Wireless Enumeration - NetStumbler



Wireless Enumeration - NetStumbler - 2

Melbourne netstumblings

<Provided during presentation>

Orinoco Tools

- If you own a WaveLAN/Orinoco card you already have enumeration tools
  - Discover nearby APs
  - Special Orinoco "ANY" ESSID/Network Name
  - (screen shot)

OK... you've found my AP

- Your AP may be **very** helpful
  - Either it or a service behind it may be providing public DHCP services
  - Potential MiM from your parking lot!
- Too much information?
  - Domain names, registered addresses

Access Risks

- Access points have defaults
  - Because of MAC-like address and service fingerprinting APs is simple.
    - http://aptools.sourceforge.net
  - Network Name, Administrative access via telnet/HTTP with default user/password - If access to AP config occurs then "Game over man. Game over."

Access Risks -2

- If WEP is enabled -matter of time...
  - Airsnort - tool to allow brute-forcing of WEP by using collected traffic
    - http://airsnort.sourceforge.net
    - Only works with PrismII chipsets
  - WEPCrack
    - http://wepcrack.sourceforge.net
    - Rough set of perl scripts

Access Risks -3

- If WEP isn't enabled - y0r 50 0wn3d
  - If AP is a bridge then traffic on the **wired** lan can be driven to a wireless client using arp/dns spoofing - just like LAN clients
  - Defense attempts like only offering services like DHCP by known MAC addresses or filtering on IP addresses are easily defeated

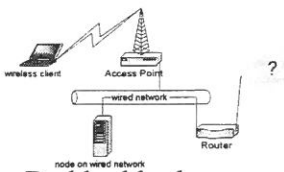
What can be done?

- THINK before implementing.
- Old AAA (authorization, access, accounting) should be our goal
- We have the technology...
- Use obscurity for secondary security

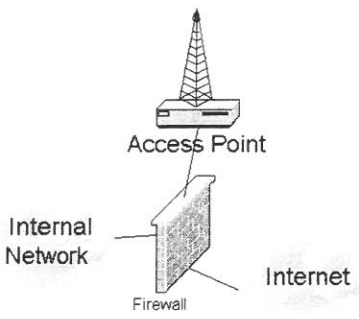
THINK before implementing

- Do you **really** need this?
- Wireless is an additional network perimeter. It should get such consideration in your network design
- Know the answer to the question "How am I going to police this service?"

THINK before implementing - 2



Bad bad bad



Segregate and regulate

AAA should be our goal

- AAA mumblemumble RADIUS mumblemumble TACACS+ mumblemumble
- Control **who** goes **where** and monitor **how** and **what** they do.
- Consider the risk of being a "stepping stone" or the source of an attack on third parties

AAA should be our goal - 2

- Find ways to account for activity - MAC addresses, AP activity. Use an IDS -or just tcpdump **everything**
- Design a method to enforce authorization for use of the service
- Limit unauthorized access by eliminating anonymous use

We have the technology...

- Utilize VPNs over 802.11
  - Gives better ability to authorize use in a vendor-independent way
  - Strong encryption!
- Control use with strict egress filtering
  - Another use for your overpowered firewall

Obscurity <u>can</u> be ok

After doing the right thing with primary security design...

- Put generic details in AP and services
  - Generic, non-disclosing ESSID names
  - IP addresses, domain names from DHCP
- Use NAT. Avoid telling attackers who you are by using RFC1918 addresses

Obscurity is usually "free" requiring only configuration to implement...

Other URLs

- http://www.wardriving.com - kit info
- Fluher, et al paper on RC4 in WEP issues:
  - http://www.eyetap.org/~rguerra/toronto2001/rc4_ksaproc.pdf
- EvilPete scripts
  - http://www.dis.org
- Dsniff
  - http://www.dachb0den/projects/bsd-airtools.html
- Local wireless efforts
  - http://melbwireless.dyndns.org -good geeky fun
  - http://www.air.net.au - Canberra (and other). Great antenna info

Thanks!

- To you for your attention
- Wardrivers - Michael Flower, Keith Glennan, Stephen Lynch, Jeff Paine, Dean White, Cath Wise

Understanding PKI Issues.

See http://www.auug.org.au/security2001

Secure Linux Workstations with filtered NFS

Matthew Pratt and Bob Edwards
Department of Computer Science
Australian National University
Canberra, Australia, 0200

14th November 2001

Contents

1 Introduction

Linux-based PCs in Student Laboratories provide a low-cost and robust environment for teaching specialised subjects such as Operating Systems and Networking, especially given that students have full access to the Linux source code.

In such environments, students often need to be able to use the workstations as the super-user. It is also advantageous if students are able to gain full access to their home directories from the central file-servers. These two requirements introduce some interesting issues, especially in the important area of file security.

In some of the laboratories at the Department of Computer Science at the Australian National University in Canberra, we wanted to allow our Linux-based PCs to mount student directories from the existing Sun Solaris file servers running Network File System (NFS). This paper discusses how we went about implementing this.

1.1 Requirements

For the system to be put into production it had to meet a number of requirements:

1. Linux PCs integrated with the existing computing environment (including acting as X-terminals when necessary)
2. Linux PCs to provide an identical environment, and access to the same resources (i.e. a home directory) from any terminal
3. Provide a good level of security for user data, the machines and the network.

2 Background

2.1 Linux

Linux is already used extensively throughout the department, is free, integrates extremely well with the existing Sun and Linux environments, runs very well on PC hardware and is very customisable with all the source code freely available. It was the natural choice for use on the client PC systems as well as any servers needed.

2.2 Hardware and Networking

A number of PCs were already in use in the student environment as lab X terminals, and the department has an existing Ethernet infrastructure (capable of 100Mb). Further details of the hardware can be found in section 6.1.

2.3 File Sharing

File sharing is central to the distributed computing environment. Since the base of the user environment is the home directory on Unix (and Linux) systems, and it was a requirement to have access to the same environment from any terminal, the users home directory must be centrally located[1] and remotely accessible. To accomplish this a file sharing protocol is needed.

2.3.1 Network File System, NFS

NFS is a network file sharing protocol, developed by Sun Microsystems in the early 1980s. NFS was created with a number of goals in mind including performance, transparent access, statelessness, and portability. To help meet these goals NFS was built on top of Sun's Remote Procedure Call protocol (Sun RPC), whose details are covered in section 2.3.2. Most implementations of NFS rely on stateless datagram packets carried over fast local area networks (e.g. UDP/IP/Ethernet) to achieve good performance. In these situations NFS can be as fast, and sometimes faster than local disk access. Alternatives exist for NFS such as AFS, CODA and SMB, but none are as mature and well supported under Solaris and Linux as NFS.

NFS seems like an ideal solution for sharing home directories, however it is showing its age, especially in the area of security. The NFS server relies on RPCs System Authentication method, trusting clients based on their IP address at the file system level (i.e. only trusted machines can mount exported file systems), and trusting the users UID/GID sent by the client at the file level (i.e. only the users whose UID/GID allow access to a particular file/directory are given access). In a public lab environment both of these security measures can be easily overcome by a malicious user.

These faults are further described in section 2.3.3, and have been the primary reason for not adopting such a system previously. The majority of this document focuses on fixing this problem.

2.3.2 Sun RPC

The Sun Remote Procedure Call (RPC) is the underlying protocol of the Network File System.

It provides the security mechanisms for NFS and so is important to understand. The layout of a RPC packet using System Authentication can be seen in figure 1. The parts used in NFS security as mention previously are the IP address and the UID/GID found in the RPC header. RPC can also be configured to use DES or Kerberos authentication (known as secure RPC), but this is not often used and not supported by Linux.

2.3.3 The Problem with NFS

The problem with the use of the standard Network File System protocol on publicly accessible networks is that its security measures are insufficient. Machines that allow root access, or other machines that are plugged into the network, can easily defeat the IP and UID trust mechanisms that NFS relies upon. The root user on an NFS client can easily defeat the UID based security methods since it has the ability to change to any UID. If an external machine were to assume the IP address of a trusted NFS client, or spoof request packets from the IP of a trusted client, then the IP based security of NFS would also be defeated.

For example consider a student lab machine on a publicly accessible network. Trust must be given to it by the NFS server, so that it can NFS mount student home directories. A malicious user could bring in a laptop and unplug a lab machine replacing it with the laptop and using the same IP address. Since the user owns the laptop, he has root access on it, and can therefore assume the UID of any user whose file he wants to access on the NFS server. The NFS server has no way of knowing that the trusted client machine has been replaced, or that the user has not been properly authenticated to use that UID.

A second situation to consider is one where root access must be given to users for them to perform their work. This may be the situation, for example, in operating system labs, where

[1] It may not be necessary for it to be stored centrally, but it is most practical.

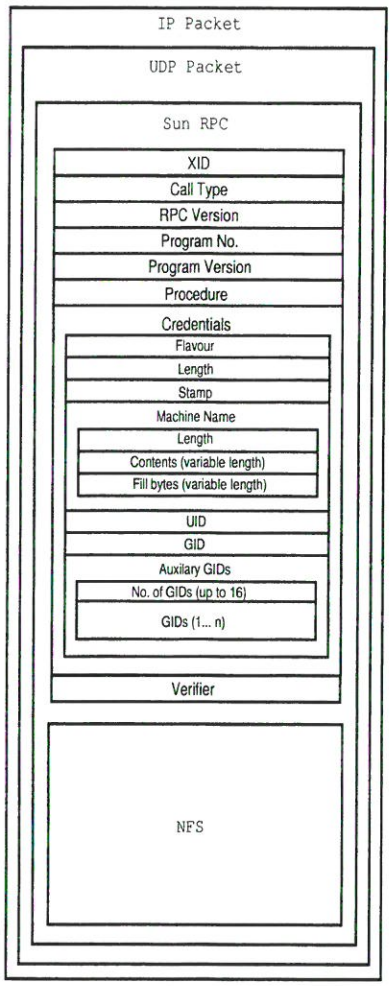


Figure 1: The NFS packet as it occurs on the wire. RPC fields shown.

root access is needed to access and replace the operating system kernel. In this situation, the user would like to still be able to access their home directories for saving of their work, but trust cannot be given to machines in which root access is allowed to all.

3 Further Requirements

As well as being able to mount home directories, we have further requirements for Linux-based PCs in our student laboratories:

- ▷ no public IP addresses - for two reasons: to prevent external attacks on potentially ill-configured machines and to prevent students from putting up their own publicly accessible servers (web, irc, ftp etc.)
- ▷ still need to retain X-terminal emulation functionality
- ▷ minimal impact on existing NFS based services
- ▷ all publicly accessible workstations to be "protected" from the Internet at large

4 Possible Solutions

A number of possible solutions to this problem were looked at:

- ▷ secure NFS - Sun Microsystem's implementation of NFS over a secure version of RPC. Each RPC packet contains a DES authentication token. Problems with this solution include a lack of client support for many operating systems. A Linux version has not been available since the 2.0 series kernels. Possible performance issues at the server. Administration issues.
- ▷ replace NFS - with a "session-based" remote file system protocol such as Microsoft's SMB protocol. Again questions of client support and maturity come into play, as well as having to move away from the current tried and proven system.
- ▷ modification of the NFS server - The server daemon could be modified in such a way that it worked with an additional authentication method. Problems of source code availability of on the existing servers then arise.
- ▷ inspection of NFS requests - A firewall that inspects the contents of NFS packets destined for the server combined with additional authentication mechanisms could be used to determine which packets should be allowed to pass. New code would have to be written to perform the filtering as well as the authentication mechanism.

5 The Solution

The solution chosen involves placing a Linux-based router between the Sun Solaris NFS server and the laboratory workstations. This router implements IP-Masquerading, otherwise known as (Source) Network Address Translation (NAT). All NFS requests from the client workstations appear, to the server, to be coming from the router's IP address. The router is placed in a secure location and has an IP address which is trusted by the NFS server. It is set up to "filter" the NFS requests from the client machines in such a way that only one valid user ID (UID) can come from each client machine, and this UID is known to the router. This filter code was written as part of the Linux kernel Netfilter structure by Matthew Pratt.

5.1 Network Address Translation (NAT)

Network address translation is a form of packet mangling. It is considered by some to be somewhat of a 'hack', but it offers a number of benefits.

5.1.1 Why do NAT?

There are a number of situations where NAT is very useful. The first is in the situation of a shortage of publicly reachable Internet addresses. A series of client machines can be given non-publicly accessible IP addresses and placed behind a NAT router as shown in figure 2. The clients will still be able to make connections to the outside network and data associated with those connections will be allowed back, but connections directly to the client machines will not be possible. This last point may be considered an advantage in some situations, since the client machines are safer from attack from the greater Internet. Another implication of this form of NAT is that itinerant servers that may find their way onto client machines are not accessible from the greater Internet.

Another situation where NAT may be useful is that of providing transparent proxying. NAT can be used to force all packets that reach a router to a particular machine, possibly with a proxy setup on it. The client machines never need know that their connections are being forced through a proxy.

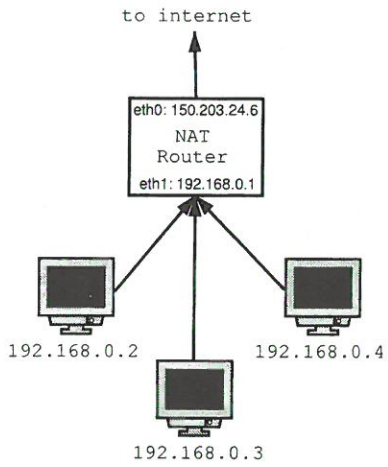


Figure 2: Useful NAT setup

5.1.2 How it works

Network address translation works by modifying either the source address or destination address of and IP packet, typically as it passes through a router. The NAT router must remember how it mangled the packet so that it can deal with any reply packets. NAT can be divided into two types: Source NAT and Destination NAT, depending on which IP address is modified. Destination NAT is useful for transparent proxying and load sharing, while Source NAT is useful for connection sharing and masquerading. Since Source NAT is primarily what is used in our situation it will be considered more closely here.

Source NAT typically occurs near the end of the packets traversal of the network. The source IP address is changed and check sums are recalculated. Since a reply will typically come to a packet, the NAT code must keep track of the source of the original packet so that it knows where to send any response. Typically this is done by the NAT code using a separate high port to send packets for each connection. The original source IP address will be associated with this port, and packets arriving at the port for the duration of the connection will be sent to this address.

5.2 Linux Netfilter

Netfilter is the network packet filtering and mangling framework introduced in the 2.4 series of kernels. An overview of packet traversal through Netfilter can be seen in figure 3. At its lowest level Netfilter is simply a set of 5 hook points that are represented by rectangles in figure 3. The

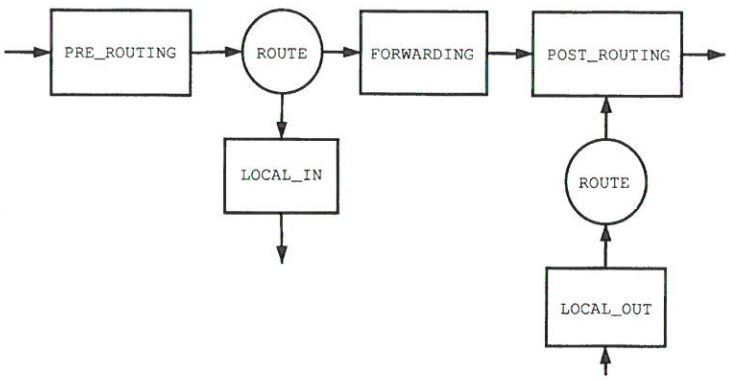


Figure 3: The Netfilter architecture

kernel will execute all the call backs hooking into a point when ever a packet arrives at that point. Each call back is free to manipulate the packet that it is passed. This means modules that extent the basic functionality of Netfilter can easily be made, since they can be made to listen at any of the 5 hook points. The module can tell Netfilter to do a number of things:

- ▷ NF_DROP: drop the packet. The packet does not continue traversing through netfilter.
- ▷ NF_ACCEPT: continue traversal as normal.
- ▷ NF_STOLEN: the module has taken the packet and will not continue traversal.
- ▷ NF_QUEUE: queue the packet for (possibly) user space handling.
- ▷ NF_REPEAT: call this hook again.

Netfilter comes with a range of modules designed to perform a range of common tasks. These include:

- ▷ Network Address Translation
- ▷ Packet selection and filtering
- ▷ Connection tracking
- ▷ Packet logging

5.2.1 Netfilter and NAT

The NAT module that comes with Netfilter hooks into the PRE_ROUTING and POST_ROUTING points. Depending on whether Source NAT, or Destination NAT is to be performed the packet is mangled at one of these points. To ensure that fragmented packets are not mishandled, Netfilter reassembles all fragments before performing any NAT operations.

5.2.2 Packet selection and user space control: IPTables

A packet selection, mangling, filtering and NAT framework has been built on top of Netfilter called IP Tables. IP Tables follows its predecessors, and counterparts on other UNIX variants in that is uses a packet traversal paradigm. Packets traverse tables of rules which specify how to deal with the packet based on properties of the packet. This makes firewall building possible since malicious packets can be detected and dropped while normal packets continue unaffected.

5.3 How it works

The solution chosen was that of packet inspection as requests are transmitted to the server. A diagram of the concept of the solution can be seen in figure 4. A number of reasons prompted this solution, including the release of the new Netfilter architecture in the Linux kernel and the benefits afforded by using Network Address Translation on the firewalling machine. The Netfilter architecture allows modules that hook into the IP stack to be easily written. Using Network Address Translation on the firewalling/packet inspection machine has a number of advantages. It means the client machines can be on a private sub net and so will not be addressable from the outside network (i.e. the Internet) which increases security for those machines. It would be impossible to run any unauthorized publicly accessible servers or services on these machines. The Network Address Translation framework in Netfilter reassembles fragmented packets so that it can track connections, and which simplifies the NFS packet inspection code, since it will never have to deal with fragmented packets.

The filter server shown in figure 4 acts as a firewalling NAT router. Each packet destined for the NFS port (2049) would be inspected by a kernel module that hooks into the PREROUTING point of Netfilter. The module would compare the source IP address of the packet as well as the

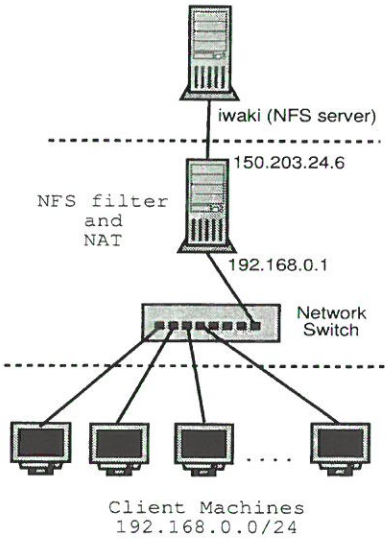


Figure 4: Network Topology of NFS filter solution

UID/GID credentials against a table of known users that is maintained by a user space daemon. This daemon only adds entries to the table when a TCP connection come in from one of the client machines, and a user authenticates his/herself over the connection. When the connection is broken the entry in the kernel table is removed. A detailed view of the filter machine can be seen in figure 5. In a sense, the filter server acts as a way of providing authenticated, session-based

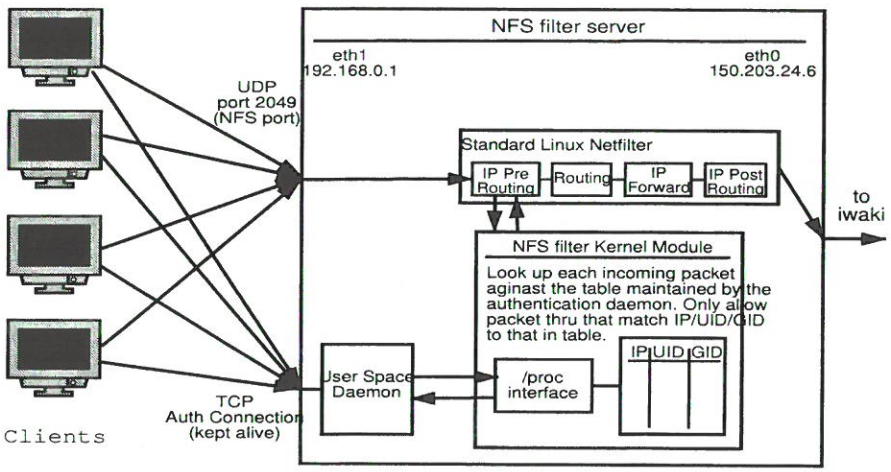


Figure 5: Filter server internals

NFS access to individual machines.

5.4 NFS filter kernel module

The NFS filter module is a kernel module that when inserted into a running kernel directs Netfilter to pass all packets at the IP prerouting stage to it. By this stage each packet has been fully assembled into a buffer (and defragmented thanks to the NAT layer), and checked for errors. The module simply passes on any packets which are not UDP/IP and are not destined for port 2049, the NFS service port.

All packets that make it past the checks are assumed to be NFS packets. The RPC header will be the first piece of data inside the packet, and is what we are interested in since it holds the

UID/GID credentials of the NFS request. If the UID/GID extracted from the packet match the UID/GID in the table with the same IP then the packet is allowed to pass untouched.

The table previously mentioned contains three values for each entry: the IP of a client machine, the UID of the user logged into that machine, and the GID of the user. The IP entry is assumed to be the unique key in the table. It is also assumed that an external mechanism (user space daemon) manages the contents of the table, and that an absent entry means that no user has been authenticated for that machine.

A `/proc` file system interface provides access to the table from user space. Two files allow manipulation of the table. `/proc/nfs_filter/table_add` provides a method of adding entries to the table. Writes to this file are expected to be in the format: <ip address> <uid> <gid>. This file can also be read to get an indication of the state of the table. `/proc/nfs_filter/table_del` provides a way of removing entries from the table. Writes to this file are expected to be the IP address of an entry to remove, or a '*' to remove all entries in the table.

5.5 Design problems and trade offs

How to deal with non-conforming requests was discovered to be a problem. That is NFS packets that are from a non-registered UID/GID. The original plan was to simply drop these packets, which is easy in the Netfilter framework, since all that is needed is that he module return NF_DROP. By the problem is that the Linux kernel NFS client implementation (and possibly others) will not issue any further NFS requests to the server until the outstanding one either succeeds or fails. As a result the client must receive some form of notification of bad request. Here we have a number of options:

- ▷ Have the NFS filter module construct a valid, 'permission denied' NFS packet and return it to the client.
- ▷ Have the NFS filter module construct a ICMP error packet (such as port unreachable) and return it to the client.
- ▷ Modify the RPC credentials of the packet to a benign value (nobody user) and pass it onto the server. The server will create a permission denied message and return it.

The first method was tried, but added much more complexity to the module, since a NFS on RPC on UDP on IP packet must be created from scratch with all options and check sums created and filled it. It was discovered easier to modify the UID/GID fields of the RPC header, recalculate the UDP check sum and allow the packet to continue on its way.

The second problem was that checking the UID/GID inside the Sun RPC header is that it is not enough. RPC provides space for an series of additional GIDs since any user may be a member of more than one group. These additional GIDs must also be checked, since a malicious user may fake membership of a group and bypass the UID check. The question then arises as to whether it is worth storing all the GIDs of each user in the kernel space table. The table must be looked up and checked against every packet NFS packet that passes through the filter. Additionally there is the problem that the auxiliary GIDs arriving in the packet may not be in the same order as those in the table, requiring either a sort to occur, or a more sophisticated comparison algorithm. All of this additional checking adds up to degraded performance, and bloat in the kernel IP/UID/GID table and module code.

The other solution is to assume the user will only ever be allowed to be in one group, the primary GID group, and force all auxiliary GIDs to this GID. Since the packet mangling code had already been developed for the first problem, it was easiest to implement the GID forcing solution. This solution may be extended in the future to force to multiple fixed GIDs instead of just one.

5.5.1 The user-space daemon nfs_filtd

The user space daemon is charged with a number of responsibilities including authenticating users, maintaining the filter module's filter table, and setup/tear down of firewalling rules for

authenticated hosts.

The primary authentication method for the system is a SSL over TCP connection from the client machine. The NFS filter and firewalling rules are only in place as long as this connection is present. This is to prevent a malicious user from replacing a client machine with another after they have been authenticated, since the TCP connection will not be established with the new machine. In effect it provides the session mechanism for the system. The nfs_filtd expects a heartbeat from the client every second. If a number of heartbeats have been lost then the daemon assumes that the machine has been disconnected or re booted and removes the filter and firewall rules for that host.

The primary function of the user space daemon is to listen for and manage the SSL/TCP session connections from all the clients. The user name and password of the user wishing to be authenticated is expected over the connection when it is established. The daemon performs authentication of the user and if successful establishes a filter table entry with the originating IP and UID/GID of the user, as well as a NAT rule with IPTables. Upon breaking of the connection the filter table entry and NAT rule are deleted.

5.5.2 Client side authentication

Each client authenticates itself against the filter server daemon at log on time. The SSL/TCP session is maintained as long as the user is logged into the machine.

Under Linux this is quite easy thanks to the use of the Pluggable Authentication Modules (PAM) system. PAM enabled applications (essentially all included with a distribution) request authentication service by name. For example the login program would ask for the 'login' authentication service. The PAM subsystem will then load the configuration for that service (typically from a file like /etc/pam.d/login). The configuration will specify which modules to use for authentication, and in what order to execute them. A further extension used by Red Hat is that all PAM configuration files refer to a central authentication stack named *system-auth*. It is therefore only necessary to modify the one configuration file to have the whole system authenticate against a new service such as the one we implement.

The first problem that arises is how to generate a heartbeat that is sent to the server every second to let it know that we are still connected. The PAM session authentication module is executed once at log in and must return for the login to succeed and so cannot contain a loop to send the heartbeat signal. One way around this is to have the login process fork() and have the child send the heartbeat. This has a number disadvantages including having a forked login process running for each login, having to detect when the user logs out, and increasing the complexity of the PAM module.

Another solution, the one used, was to have an authentication daemon (called the nfs_authd) running on the client that manages the sessions. The PAM module connects to the daemon (which is started at boot) via UNIX named pipes (FIFOs) whenever authentication is required. It sends its process ID (PID), the user name and password before it disconnects. The daemon connects to the nfs_filtd daemon on the server using SSL and attempts to authenticate. If successful the nfs_authd mounts the users home directory and replies to the PAM modules that authentication was successful. It checks that the initiating login process is still alive (it has its PID) every second, and if it is sends a heartbeat to the server. This is done for each current login. Once the login process dies, due to the user logging out, or for any other reason (like an Xserver crash), it tears down the SSL connection so the server.

The method of monitoring the PID of the initiating login process was used in place of using the session mechanisms offered by PAM to prevent the PAM module having to reconnect to the nfs_authd at log out, and to ensure that sessions are stopped when the login process dies unexpectedly and the PAM session close procedure is not run.

Another problem was discovered with the authentication method. All login mechanisms check that the user is valid and has a valid UID/GID before proceeding to the PAM authentication stage. Typically this information would come from a NIS+ service provided to the machine. Since the client machines are firewalled from the rest of the network until an authenticated user is logged

in, then cannot do a look up against NIS+ or similar. The solution used was to install a LDAP that contains all student UNIX account information, and allow look ups against this server at all times. A LDAP Network Switch Server module on the client queries this server when needed.

To handle mounting and unmounting of student directories, the nfs_authd executes shell scripts. This is done to allow the administrator to easily change the configuration, without touching the code of the daemon or the PAM module.

6 Complete implementation

6.1 Hardware

Client hardware consists of a Intel Celeron based PC, with 128Mb of RAM, 8Gb hard disk, TNT video, and Intel EEPro100 Ethernet. Server machines consisted of similar hardware with an added Ethernet interface (a DEC Tulip compatible chip set[2]).

6.2 Network topology

The network topology for the ANU DCS linuxlab environment can be seen in figure 6. All links are 100Mb Ethernet. It was decided to run most of the services on a single machine on the private

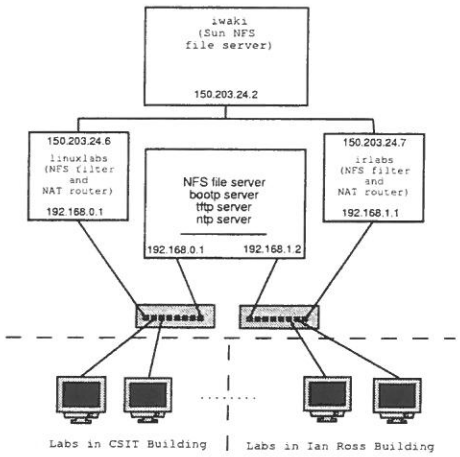


Figure 6: Network topology of student Linux lab environment

sub net, with no access to the external network for a number of reasons. The first is for ease of administration. Installing packages on this one machine would make them available to all the clients. The seconds reason is for security. The machine would be running services with a spotted security vulnerability history. Being only accessible on the private sub net means that it is hidden from the wider Internet, and it it were to be compromised, would have no external connection.

Since the entire client system comes off this server, it provides the central administration advantage typically given by a traditional central computer server.

6.3 Netboot

To be sure that the workstations remain secure and that none of the important software has been modified or trojaned, the workstations boot a copy of Linux off of the network. Upon start up each machine is setup to make a broadcast bootp request (using etherboot). This request is serviced by the main private sub net server as shown in figure 6. The bootp response contains the assigned IP

[2]It was found that the tulip network driver supplied with the 2.4 series Linux kernels did not function satisfactorily with the cards we had. The de4x5 driver supplied with the kernel proved to be successful.

address as well as the default gateway, and a server address for where to get the boot image from via tftp. The boot image consists of a 16MB ram disk that was built up using a working version based on Red Hat 6.2 to a 7.0 version. The image contains a minimal root file system including /etc, /bin, /sbin and /lib and is configured to mount the /usr file system read-only via NFS off of the same server and start a X session with appropriate login/chooser widget. As described the bootp, tftp and /usr NFS server is located on both private sub nets for speed and security reasons. This means that the only NFS packets going through the NFS filter boxes are those for user home directories and must be monitored.

6.4 Xterminal Use

One of the requirements for the machines behind the NFS filtering firewalls was to also act as Xterminals so that users could log directly into the Sun servers when necessary. Ideally the user would be able to choose which machine to log into whether local or remote, and login time. However the X and XDMCP (X display manager control protocol) protocols were not designed to be used with NAT, and provided some trouble.

6.4.1 XDMCP and X vs NAT

The X protocol reverses the traditional client/server model we have been using. With X, the server is run on remote machine (which we have been calling the client until now), while the application is considered the client (which we have been calling the server). The X software suite also provides a program called XDM (X display manager) for managing both local and remote X sessions and logins. The XDMCP protocol used by XDM is UDP based, and is typically run on port 177.

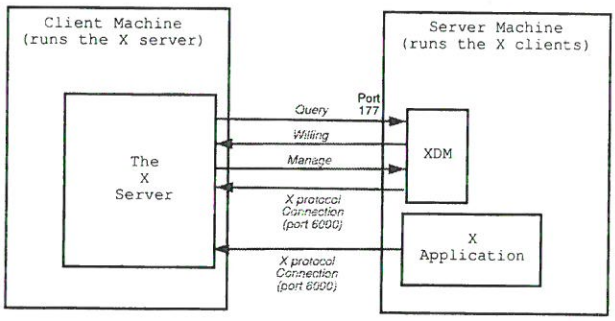


Figure 7: An example XDM session

A typical XDM session can be seen in figure 7. When directed to do so, the Xserver on the client machine will query a server running XDM. If the server is able to take logins then XDM will respond with a packet saying that it is 'willing' (typically with other information such as load etc). The Xserver will then request to to be managed by the server, by sending the 'manage' request with the machines IP address, and the display number. XDM, upon receiving a 'manage' request will then connect to the X server and display a login box, where the user can type user name and password. Once authentication is successful the normal X clients apps (such as a desktop environment like CDE or KDE) are executed and connect to the X server.

The whole XDM mechanism breaks when used from behind a SNAT firewall for a number of reasons. The XDMCP 'manage' packet contains the IP address of the host sending the request. But a NAT client host will send its private sub net IP address since that is all it is aware of. The server will try to connect to this address but since it is on a private sub net it does not know where to send the following X connection.

The obvious solution would be to modify the address in the manage packet to be that of the NAT machine as it passed through. This does not solve the problem however, as a direct connection still cannot be made to machines behind the NAT router. This is because the NAT box will not be able to determine if a packet it receives addressed to itself, is actually intended

for one of the NAT machines, since it had not previously setup the port-to-connection mapping system normally used. A solution to this problem would to modify the display number in the 'manage' packet as well. The display number is used to determine which port to connect to on the X server. Display 0 is port 6000, display 1 is port 6001 etc. The NAT machine could then forward any connection above port 6000 to a corresponding client machine. e.g. port 6001 get forwarded to 192.168.0.11:6000, port 6002 -> 192.168.0.12:6000... and so on.

A second and much easier to implement solution was devised: the NAT rules could be bypassed for X and XDMCP traffic and a static route given to all servers so that they know how to get packets back to the client machines. This solution needed no new code or modification to the clients, simply some additional firewalling rules to the filtering/NAT machine, and only reduced security slightly and so was used. The firewall rules were simply inserted in the NFS filter user space daemon start up script before the NAT rules, and allow any traffic bound for port 177 UDP out and any traffic bound for port 6000 TCP or from port 177 UDP in.

6.4.2 XDM and login host selection

The XDMCP protocol includes a method for host selection. An 'indirect' query can be sent to a XDM server which will forward that query to all known XDM servers and present a chooser list to the originating X server. When the user selects a host the standard process shown in figure 7 takes place. To allow the user to log into the local host we must therefore run a local XDM server and which is aware of all of the other XDM servers and tell the X server to do an 'indirect' query or the local XDM server.

The X server is launched by init like a tty so that it can respawn when killed. A typical inittab line for launching X would be as follows:

```
x:5:respawn:/usr/X11R6/bin/X -indirect local_host
```

The X server sends the IP address of itself in the XDMCP 'manage' packet and which it determines from its command line arguments. Because of this the 'localhost' name cannot be used since this resolves to 127.0.0.1 which will be sent in the packet and will not be able to be used by XDM. As a result a special entry need be added to the /etc/hosts file at boot to identify the local IP address. This is done in the `/etc/rc.d/rc.sysinit` file as follows:

```
ADDR=`/sbin/ifconfig eth0 | grep 'inet ' | awk '{print $2}'| cut -f2 -d":"`
echo "$ADDR local_host" >> /etc/hosts
```

The local XDM daemon must me directed to allow connections from the local machine, and to pass indirect queries onto the other servers. This is done via `/etc/X11/xdm/Xaccess` as follows:

```
# Allow the local machine to connect
local_host
# Indirect to the other servers as well as self
%hostlist iwaki local_host mehta
* CHOOSER %hostlist
```

6.5 External Access

One of the expected benefits of having all the client machines on a private sub net behind a NAT firewall is that they cannot be accessed directly from outside the private sub net. However this benefit can be restricting. It became clear that exactly this functionality would be needed by some clients. Some of the client machines would need to be used as student run servers. Fortunately, using Destination NAT the problem can be solved. The solution involves giving the NAT machine another IP address on its external interface (eth0) for each client machine that is to be accessible. Iptables is then setup to Destination NAT any packets coming in on that interface to the corresponding client machine on the private sub net. Iptables is also used to ensure that only selected ports are allowed through (namely those that are considered 'safe' like ssh, www and https).

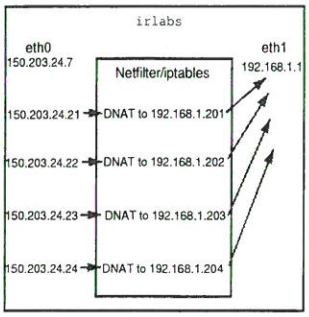


Figure 8: The DNAT configuration used for external access

6.6 The dual LDAP server system

Traditionally the systems in DCS have used a NIS+ feed from TLTSU for authentication and user account information services. However, there is a migration away from this system for a number of reasons including an industry shift toward LDAP (including a shift by Sun), poor NIS+ support under Linux, the availability of a central campus LDAP server, and the problems of password propagation within the feed. All of these reasons prompted our shift to LDAP.

There was one stumbling point however. The campus LDAP server contained none of the Unix account information that is needed for a complete authentication service. There were a number of solution possible:

- ▷ convince the campus LDAP maintainers to include Unix account information in their database
- ▷ maintain our own local LDAP server that contains Unix account information but replicates the rest from the campus (password propagation problems would exist)
- ▷ modify our local LDAP server to pass authentication requests to the campus LDAP server (requires source code to LDAP server)
- ▷ run a local LDAP server that contains Unix account information, and modify our clients to authenticate against one server while getting account information from another

Fortunately the Pluggable Authentication Module system used by Sun and by most Linux distributions makes the last solution the easiest to implement. As a result a custom PAM module was written for our Sun Solaris systems and the nfs_filtd was modified use this system.

Another problem found with the system at a later stage was that only user accounts exist in the campus LDAP server, meaning any local accounts, such as group accounts were no easily accessible. The solution was to store the local account in the local LDAP server and have all authentication modules fall back to trying to authenticate against the local LDAP server should authentication against the campus server fail.

6.7 nfs_filtd10

The server user space daemon's first implementation was as a pure TCP event driven daemon, using selects and call backs according to what input was received. As work progressed and SSL support was added, it was discovered that OpenSSL could enter a state where it would block for a length of time. Likewise, should a delay occur on any LDAP look ups, new users would not be able to authenticate, heartbeats may be lost and users disconnected. To make the daemon more robust to these sort of problems it was moved to a forking based model, where the daemon forks for each connection. This model brings its own problems, mainly how to enforce connection limits. Each server child process must be aware of the details of other connections. Information such as the number of connections from a particular IP (so only a limited number of connections will be accepted from a single client, preventing a Denial of Service attack), whether another user

is already logged in on an IP, and how many connection there are overall. This answer to this problem was to provide a shared memory area to all the child servlets that contained a connection state table. This table is locked by a posix semaphore to prevent concurrent access from multiple servlets.

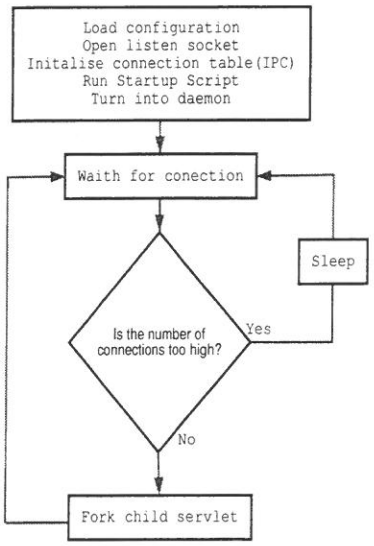


Figure 9: The nfs_filtd main process

The server can be broken down into to parts. The main daemon process, shown in figure 9, always runs and listens for new connections. It has the task of deciding if there are to many open connections and refusing any more. In normal operating circumstances it will simply fork itself and execute the servlet code that deals with the connection. The servlet process, shown in figure 10, manages the connection, handles authentication of the user, building of filter rules and monitoring of the heartbeat signal. Should the connection be lost or closed, or the heartbeat no arrive for some amount of time, then the servlet process will simply exit, taking down any filter rules it had built.

6.7.1 UID look ups and user authentication

The nfs_filtd performs UID/GID look ups using the standard getpwnam() library call. As a result a LDAP Name Service Switch module must be installed for this to work against the local LDAP server that store account information (see section 6.6). This module is supplied with most distributions and can be configured at install time. User authentication however, can not occur through this interface since password information is stored in the campus LDAP server. To perform authentication the nfs_filtd attempts a LDAP bind to the campus server. Should this fail it will fall back to attempting to bind to the local LDAP server, so local accounts will work.

6.7.2 Filter construction and destruction

To make the nfs_filtd more flexible for the system administrator IPTables filter construction and destruction was moved into a number of shell scripts. The script `/etc/nfs_filtd/new_connection.sh` is run whenever a user has been authenticated. It is passed the IP of the client machine, and builds a IPTables SNAT filter for that client machine. Likewise the script `/etc/nfs_filtd/break_connection.sh` is run whenever a user disconnects from the server and removes the corresponding IPTables rule. NFS filter rules are not added or removed from the scripts but are done by the daemon itself.

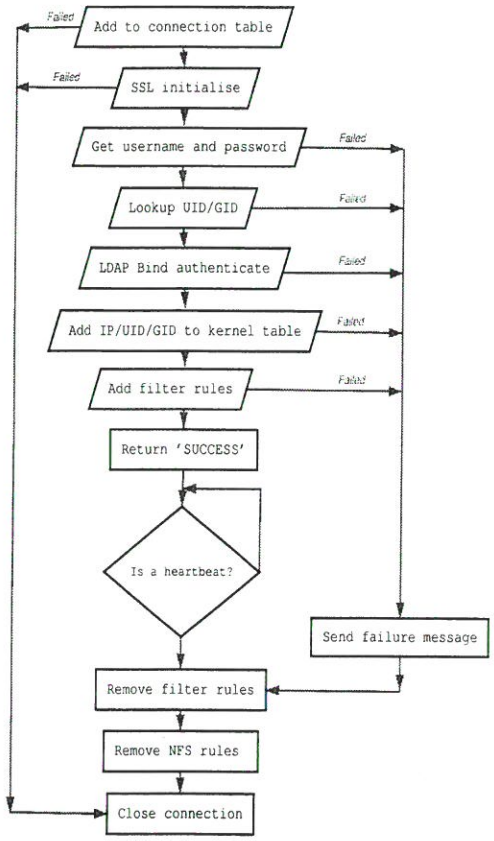


Figure 10: The nfs_filtd servlet sub process

6.7.3 Daemon Start up and Crash protection

In the case of daemon crash (due to a bug, or malicious attack), the filter table and firewalling rules may be left in an undefined state. In this case, unauthorized packets may be able to make it onto the outside network. To prevent this happening, NFS filter user space daemon listens for a range of signals that indicate shutdown, or failure including SIGSEGV, SIGTERM, and SIGKILL. Should any of the signals occur, a script, `/etc/nfs_filtd/die_script.sh` will be run. This script is responsible for ensuring no damage can occur, by removing all NAT rules, removing all NFS filter table entries, and by disabling IP forwarding.

To ensure that the correct rules are instated on start up another script, `/etc/nfs_filtd/start_script.sh` is run when the daemon is started. This script builds rules for services that need to be enabled full time such as SNAT for LDAP look ups (port 389), routing for XDMCP packets (port 177) and blocking all other traffic until a user is authenticated on a machine.

6.7.4 Configuration

Configuration for the nfs_filtd is done from two places: the run-time loaded configuration file, and compile time header file. The configuration file is stored in `/etc/nfs_filtd/nfs_fild.conf` and contains a number of adjustable options as well as comments describing those options. It is intended that the configuration file be self documenting, and so complete coverage will not be given here. The main configuration option in the file are the LDAP servers used for the authentication of users. nfs_filtd will attempt to authenticate against all the servers supplied in order. This is done so that authentication fall back described in section 6.6 is possible. Typically the campus LDAP server is specified at the main authentication LDAP server and the local LDAP server is specified

as the auxiliary LDAP authentication server. The default configuration file can be generated from the daemon itself by running it with the "-P" option. The location of the configuration file can also be specified at run time with the "-c" command line switch.

The header file `daemon.h` specifies a number of compile time values including the default port, the maximum number of concurrent connections, the location of the configuration file and scripts, the location of the daemons SSL certificate, the debug log file, and the maximum lost heartbeat time.

6.7.5 Debug mode and logging

If the preprocessor symbol DEBUG_FILE is defined in the daemons header file, then debug information will be logged to the file pointed to by that symbol. Additionally the debug information can be seen on the console when the nfs_filtd is started in non-daemon mode with the "-d" command line switch.

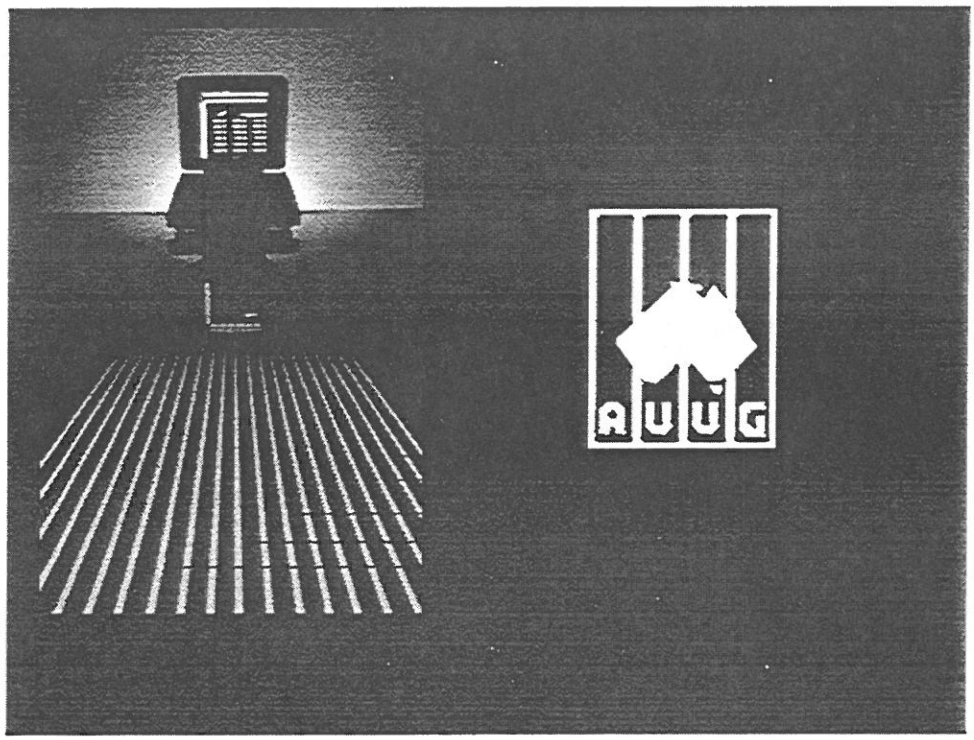
6.8 nfs_authd and pam_nfs_filter

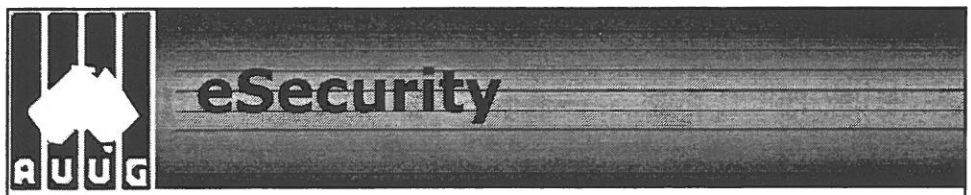
The nfs_authd daemon and pam_nfs_filter module manage the client side of the system. The PAM modules is a simple module that collects the user name and password from the user or programs and connects to the nfs_authd. The user name and password are send to the daemon along with the PID of the calling process (such as login). The daemon will return the result of the authentication request, and the PAM module will allow or disallow the login accordingly.

The nfs_authd listens for requests on a named Unix pipe (FIFO), and writes the results of any request to another FIFO. When a request comes in a connection is made to the nfs_filtd and the user name and password are forwarded over the connection for authentication. Should the authentication succeed, a script is run to mount the users home directory, and the connection is kept open. Every second the nfs_authd daemon checks that all the processes associated with its connections (i.e. those PIDs that were passed in with the user name and password) are still running. If the are then a heartbeat byte (the "\n" character) is sent of the appropriate connection. Should the process no longer exist a script is run that kills the users processes and unmount their home directory, and the connection is closed.

6.8.1 Problems with client side caching

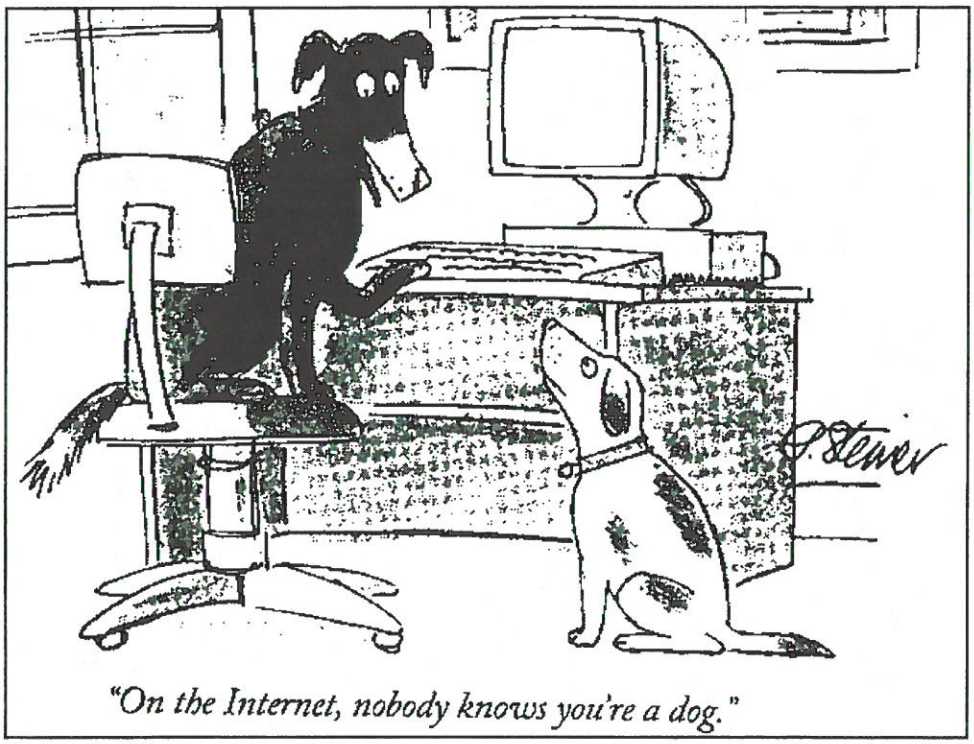
The Linux kernel caches file accesses for all file systems using the page cache and the VFS layer. This introduces a problem with security in our lab environment since the contents of a users home directory may linger in memory after the user has logged out. If the next user logs in as root and changes his UID/GID to that of a previous use, then he may be able to get cached versions of the previous users files. A simple but not complete solution to the problem is to mount each users home directory at login and unmount it at log out. Cached files may still reside in memory but the kernel will not grant access to them once the file system has been unmounted. The problem with this approach is that it adds significant complexity to the login process. For example to unmount a users home directory not file may be open on that file system. To ensure this is true, all of the processes of that user must be killed before the unmount is attempted. Additionally the unmount must occur before the filter rules have been removed on the server, else the unmount will hang.



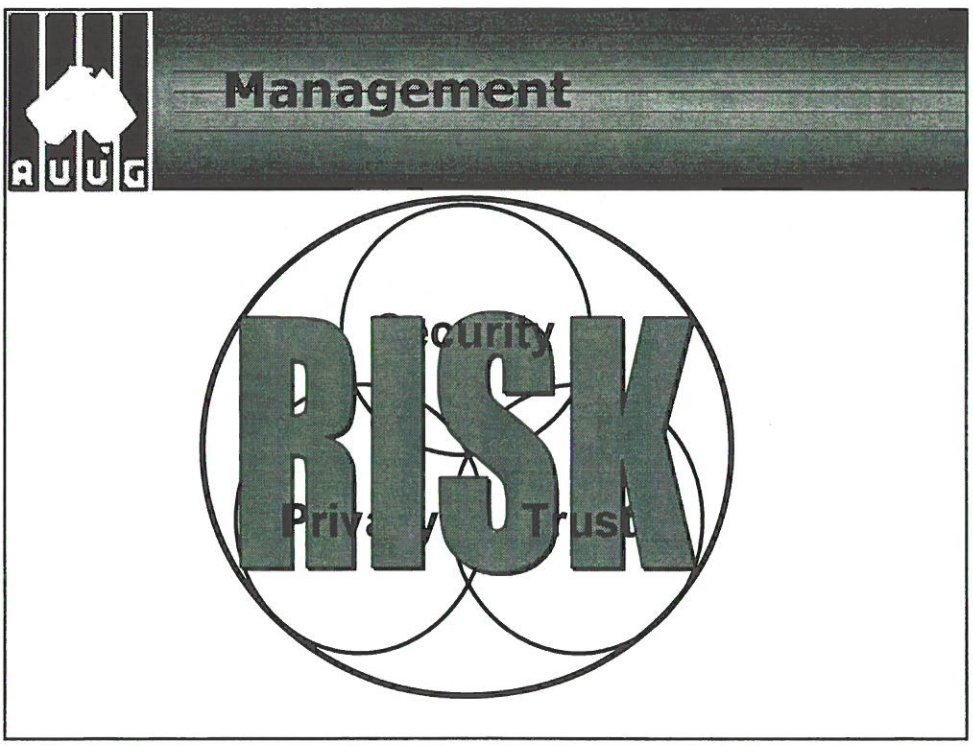


"In the brave new world of the Internet, security is an enabler of functions and activities that would otherwise be impossible."

-- GIGA Information Group, 1998

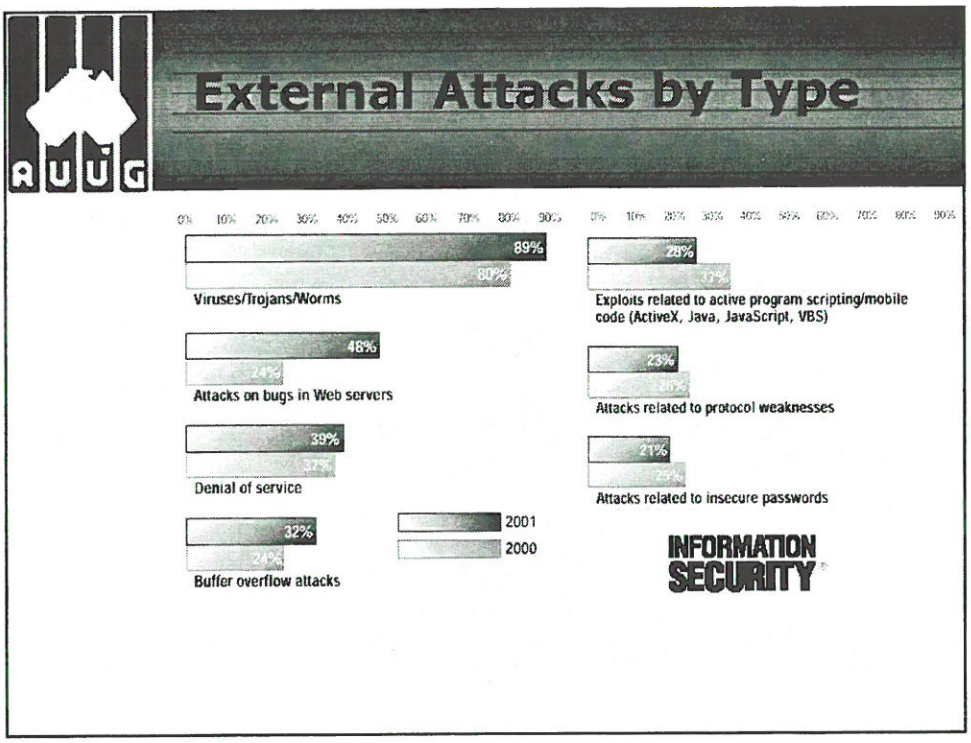


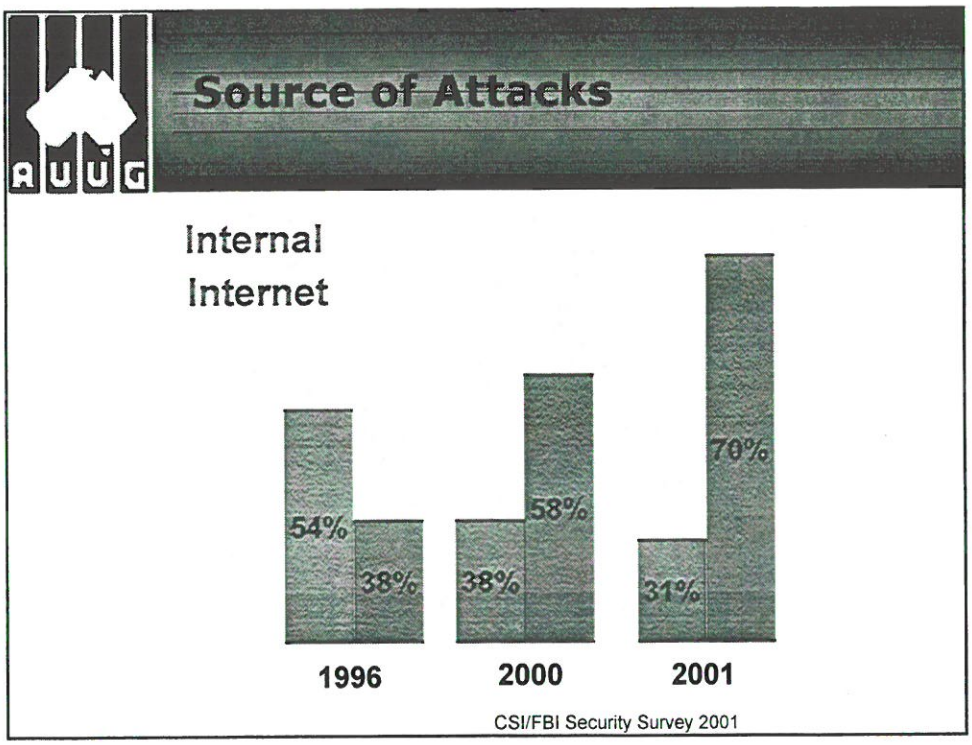
"On the Internet, nobody knows you're a dog."



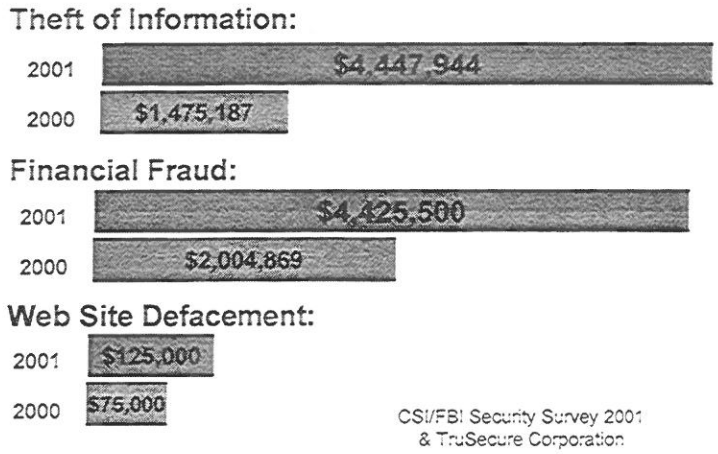
The Threats

The Threats


External Attacks by Type
89%
80%
Viruses/Trojans/Worms
48%
24%
Attacks on bugs in Web servers
39%
37%
Denial of service
32%
24%
Buffer overflow attacks
2001
2000
28%
Exploits related to active program scripting/mobile code (ActiveX, Java, JavaScript, VBS)
23%
Attacks related to protocol weaknesses
21%
Attacks related to insecure passwords
INFORMATION SECURITY


Source of Attacks
Internal
Internet
54%
38%
38%
58%
31%
70%
1996
2000
2001
CSI/FBI Security Survey 2001

Average Loss per Reported Incident



CSI/FBI Security Survey 2001
& TruSecure Corporation

Business Risk Analysis

- **Based on Business Value.**
  - **What's important to *my* business?**
- **Methodology.**
  - **Core revenue streams.**
  - **Business values.**
  - **Customer/Market values**
- **Australian & International Standards.**
  - **IEC 60300-3-9 *Risk Analysis of Technological Systems.***
  - **AS/NZS 4360 *Risk Analysis.***
  - **ISO/IEC 17799:2000 *Information Security Management.***
    - **AS/BS 7799**
  - **ACIS 33 *Handbook 2, Risk Management***

Trends in CNA

- 2001 CSI/FBI Computer Crime & Security Survey
  - 70% cited their Internet connection as a frequent point of attack (59% 2000)
  - 40% detected system penetration from the outside (25% in 2000)
  - 49% cited unauthorised access by insiders (71% in 2000)

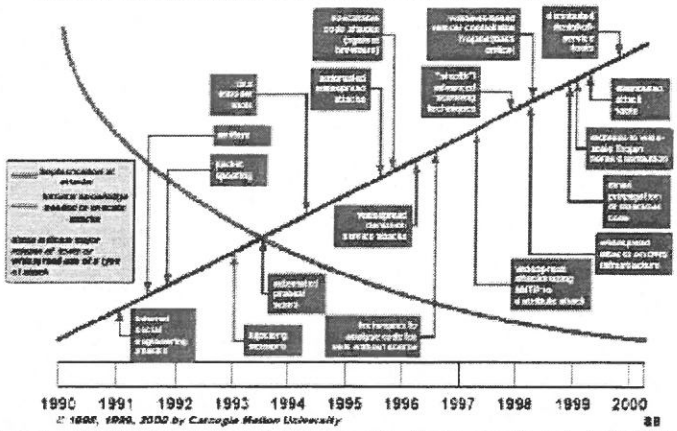
Trends

- Increasing connectivity
  - technology perceived to be affordable and beneficial by both consumers, government and business
  - about 3 million new machines connected to the Internet each month
- Growing number and skill of hackers
  - learning from younger age
  - organised and share information

Carnegie Mellon
Software Engineering Institute

CERT
Coordination
Center

Attack Sophistication vs. Required Intruder Knowledge



Honeynet Project

- http://project.honeynet.org
- consists of 8 IP addresses, ISDN connection to ISP
- uses common OS in default configuration
- no attempt made to lure attackers
  - a home network of little obvious value
- has no 'real' purpose or traffic - all activity recorded is suspiscious
  - eliminates false positives and false negatives

Honeynet Project

http://project.honeynet.org

- Lessons learnt
  - you will be attacked and often
    - » worse for advertised networks of value
  - default configurations of OS are insecure
    - » 7 default Red Hat 6.2 servers compromised within 3 days (Red Hat life expectancy 72 hours)
    - » fastest compromise time was 15 minutes
    - » default Win98, with sharing enabled, was compromised in < 24 hours and compromised 4 more times over next 3 days

Honeynet Project

http://project.honeynet.org

- attack rates are increasing over time
  - in May 2000, IDS recorded 157 alerts
    - » in February 2001, 1,398 alerts
  - in May 2000, the firewall logged 103 unique scans
    - » in February 2001, 206 unique scans
- some attackers are indiscriminate in their attacks
  - launch exploits without checking if the system is compatible or vulnerable to the exploit (ie right operating system and version)
  - allows them to scan and exploit systems in less time
- a correlation between scanning activity and successful attacks

		1999	2000	May-00	Dec-00	Jan-01	Mar-01
		Melissa	**iLoveYou**	**Kakworm**	**Hybris**	**Ramen**	**Magistr**
Infects by	User opening/running email attachment	yes	yes	yes	yes		yes
	Macros in MSWord files	yes					
	Network File Shares						yes
	by exploiting vulnerabilities			yes		yes - 3	
	Email automated (without opening) - ActiveX			yes			
	Browser automated via ActiveX/						
	javascript on web site			yes			
Propagates by	via email - attaches itself to intercepted emails				yes		
	via IRC		yes				
	via email using Address book addresses	yes	yes				yes
	via email using names in web cache						
	scans the Internet looking for computers with known vulnerability					yes	
Polymorphic					yes		yes
Payload	Easy hard availability		yes		yes	yes	
	Corrupts files					yes	yes
	Releases or intercepts files		yes		yes		yes
	installs other hacking tools		yes				
	eliminates some host-based perimeter protection						
	Installs backdoor or is able to perform root compromise					yes	
Fixes	Adjust scripting configurations or mail filtering content		yes		yes		
	Eduate users not to open/execute attachments	yes	yes	yes	yes		yes
	Keep anti-virus software up to date	yes	yes	yes	yes		yes
	Patch vulnerabilities			yes		yes	
	Disable ActiveX/Javascript						
	Disable macros	yes					
	Number of features (infection + propagation + payload)	3	6	4	5	5	6

		Mar-01	May-01	Mar-01	Jul-01	Aug-01	Sep-01
		LI0n	**Sadmind**	**Adore**	**CodeRed**	**CodeRed 2**	**Nimda**
Infects by	User opening/running email attachment						yes
	Macros in MSWord files						
	Network File Shares						yes
	by exploiting vulnerabilities	yes - 1	yes - 2	yes - 4	yes - 1	yes - 2	yes - 3
	Email automated (without opening) - ActiveX						yes
	Browser automated via ActiveX/						
	javascript on web site						yes
Propagates by	via email - attaches itself to intercepted emails						
	via IRC						
	via email using Address book addresses						yes
	via email using names in web cache						yes
	scans the Internet looking for computers with known vulnerability	yes	yes	yes	yes	yes	yes
Polymorphic							
Payload	Easy hard availability			yes	yes	yes	
	Corrupts files	yes	yes		yes	yes	
	Releases or intercepts files	yes		yes			
	installs other hacking tools	yes					
	eliminates some host-based perimeter protection	yes					
	Installs backdoor or is able to perform root compromise	yes	yes	yes		yes	yes
Fixes	Adjust scripting configurations or mail filtering content						
	Eduate users not to open/execute attachments						yes
	Keep anti-virus software up to date						yes
	Patch vulnerabilities	yes	yes	yes	yes	yes	yes
	Disable ActiveX/Javascript						yes
	Disable macros						
	Number of features (infection + propagation + payload)	7	4	5	4	5	9

Ability to Detect Attacks

- Professional penetration teams report they can penetrate 80 - 90 % of networks with firewalls and IDS with minimal chance of detection
- some hacker tools incorporate encryption, 'clean-up' features to evade detection
  - » root-kits
  - » DDOS tools
- increased sophistication of tools means that lower skilled attackers more likely to launch attacks with minimal chance of detection
- widespread infection by worms such as Nimda, Sadmind and others means that these machines are vulnerable to future confidentiality attacks
  - » by use of backdoors

Ability to Protect and Respond

- More attacks occurring
- implementation vuls increasing
  - increasing functionality/complexity results in more bugs
    - » In 2000 AusCERT reported on 421 vuls
    - » In 2001 (up to 31.8.01) AusCERT reported on 415 vuls
  - just-in-time marketing

Ability to Protect and Respond

- managing networks increasingly complex
  - » challenge to keep up with changes to the network
  - » challenge to understand how to implement and configure new services, including security features
  - » challenge to meet users demands for greater functionality and raise their awareness
    - risk assumed by one is shared by all
  - » challenge to treat security as a business critical enabler rather than as a cost centre with no return
    - Security policies and procedures must be clear, comprehensive, resourced and supported by everyone

Reaching Your Security Nirvana

- Achieving good network security is a whole of business problem – not just a technical problem
  - Defence in depth
  - Organisation wide commitment
  - Comprehensive and practical security policies and procedures
  - Checks and audits
  - Adequate resources and planning
  - Configuration management
  - Keeping up todate with threats and vulnerabilities

Great Truths

- If connected to the Internet you will be targeted often
- Default configurations of operating systems are insecure
- Hackers and their tools are becoming more sophisticated ... so must we
- Good security is a business enabler
  - Achieving good security requires a whole of business approach – don't just focus on technical solutions

AusCERT Services

Security advisories and mitigation advice for members

Confidential 24 hour incident response advice and coordination, nationally and internationally

Consultancy services

Training and education courses on computer security, incident handling etc.

Anonymous ftp:

`ftp://ftp.auscert.org.au/pub/`

World Wide Web:

`http://www.auscert.org.au/`

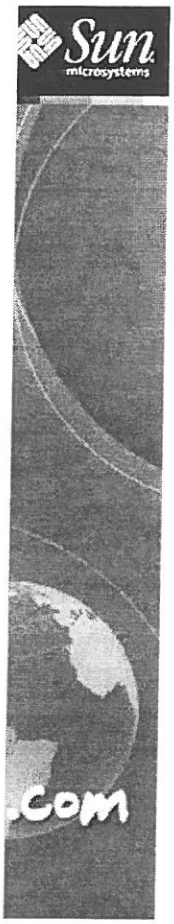
AusCERT Contact Information

24 Hour Hotline: (07) 3365 4417
(After Hours for Emergencies)
International: +61 7 3365 4417 (GMT+1000)

Facsimile: (07) 3365 7031
International: +61 7 3365 7031

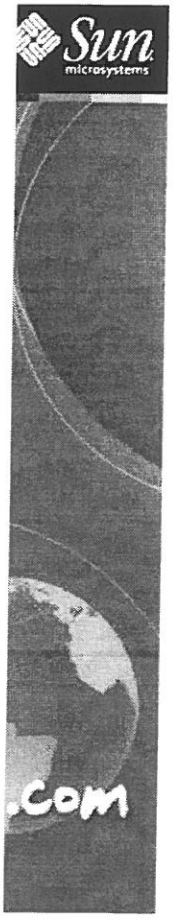
Electronic Mail: auscert@auscert.org.au
Anonymous ftp: ftp://ftp.auscert.org.au/pub/

World Wide Web: http://www.auscert.org.au/
Postal: AusCERT
The University of Queensland
BRISBANE QLD 4072



Solaris(tm) Security Toolkit Overview

Eric Halil
Network Security Engineer
SunIT Network Security Group



Agenda

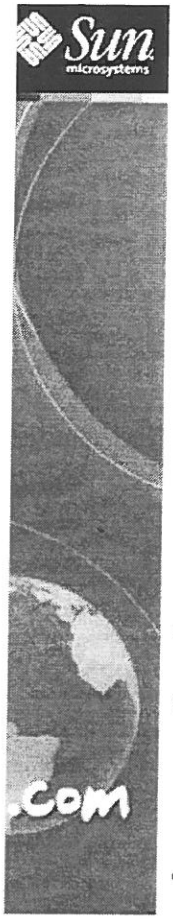
- Introduction
- History
- Approach
- Structure
- Examples
- Additional Information

Introduction

- Codification of the security recommendations as documented by the Sun BluePrints Program.
- Collection of Bourne shell scripts used to improve the security of the Solaris Operating Environment.
- Flexible and extensible framework for rapidly hardening platforms in accordance with a site's system security policy.
- Mechanism for hardening platforms in a repeatable, reliable manner for one system or 1000.

History

- Solaris is a general purpose OS that must be customised for specific installation and security requirements
- Provide a mechanism to automate making security changes to the Solaris Operating Environment and allow easy maintainability



History

- Solaris Security Toolkit (Solaris ST), informally known as "jass":
  - Version 0.1 released August 2000
  - Version 0.2 released November 2000
  - Version 0.3 released June 2001
  - Version 0.3.2 released November 2001
- Supports Solaris 2.5.1 through Solaris 8MU6
- Implements the recommendations as documented in the Solaris Operating Environment Security BluePrint article
- Not officially supported

History

- Originally developed to harden JumpStart installations.
- Consolidation of work done by many people over the last six years.
- Added command-line support and undo capabilities.
- Includes repository that provides:
  - Description of each execution
  - Installation log
  - Manifest of files copied and scripts executed

Additional Information

- Other Sun Security Sites:
  - http://www.sun.com/security
  - http://www.sun.com/software/solaris/ds/ds-security/
  - http://java.sun.com/security
  - http://www.sun.com/software/solaris/trustedsolaris/
  - http://www.sun.com/software/securenet/
  - http://www.sun.com/connectivity/suncryptoaccel1/
  - http://sunsolve.Sun.COM/pub-cgi/show.pl?target=security/sec

Acknowledgements

- Thanks to the Solaris ST team!
  - Alex Noordergraaf
  - Glenn Brunette

Australian Security and Legal Issues.

See http://www.auug.org.au/security2001